Polarization consistent basis sets. II. Estimating the Kohn–Sham basis set limit

Frank Jensen[a)]
Department of Chemistry, University of Southern Denmark, DK-5230 Odense, Denmark

(Received 6 December 2001; accepted 6 February 2002)

The performance of the previously proposed polarization consistent basis sets is analyzed at the Hartree–Fock and density functional levels of theory, and it is shown that each step up in basis set quality decreases the error relative to the infinite basis set limit by approximately an order of magnitude. For the largest pc-4 basis set the relative energy error is approximately 10^{-7}, and extrapolation further improves the results by approximately a factor of 2. This provides total atomization energies for molecules with an accuracy of better than 0.01 kJ/mol per atom. The performance of many popular basis sets is evaluated based on 95 atomization energies, 42 ionization potentials and 10 molecular relative energies, and it is shown that the pc-*n* basis sets in all cases provides better accuracy for a similar or a smaller number of basis functions. © *2002 American Institute of Physics.* [DOI: 10.1063/1.1465405]

INTRODUCTION

Density functional theory (DFT)[1] has become a popular tool for electronic structure calculations in recent years due to its favorable combination of low computational cost and good accuracy for the calculated results. In analogy with wave mechanics methods, there are two main parameters controlling the accuracy of the results, the inherent approximations in the Hamiltonian and the size of the basis set used for expanding the Kohn–Sham (KS) orbitals (here we neglect relativistic effects which become important for systems with atoms from the lower part of the periodic table). In wave mechanics the Hartree–Fock (HF) method provides the common reference, and various methods are available for calculating the remaining correlation energy. Based on theoretical analysis there are well-defined procedures for improving the Hamiltonian toward the exact nonrelativistic limit, of which coupled cluster methods currently appear to be the most popular choice.[2] The correlation consistent basis sets developed by Dunning and co-workers[3] have proven to be a good choice for systematically approaching the basis set limit for the correlation energy. Coupled with extrapolation procedures, such methods have been able to provide results of an accuracy rivaling experiments for certain properties.[4]

The main problem with DFT methods is the lack of a well-defined method for systematically improving the Hamiltonian toward the exact limit. Within KS-theory this corresponds to choosing the exchange-correlation energy functional. The local spin density approximation (LSDA) provides a reference point within DFT, similar to the HF model in wave mechanics. The introduction of generalized gradient approximations provided a large step forward in terms of accuracy, and many different functionals have been proposed.[5–9] A further improvement was achieved by mixing in part of the (exact) HF exchange energy, as first suggested by Becke.[10] Such hybrid methods are capable of giving impressive results, even for systems that are difficult to describe with wave mechanics methods. The search for functionals capable of improving the results provided by hybrid methods is currently an active area of research,[11–15] but limited success has been achieved so far. Some of these methods include a small number of empirical parameters that are chosen based on fitting to experimental data.

The other user defined component of a DFT calculation is the basis set used for expanding the KS-orbitals, but this has received relatively little attention. For applications to extended systems a plane-wave basis is often used, although recently this has also been used for smaller molecules,[16] while for molecular systems a Gaussian type basis set is commonly used. There is a general agreement that the basis set convergence of KS methods is relatively fast, and very similar to that of the HF method. The correlation consistent basis sets developed for correlation energy have been used for approaching the KS-limit,[17] but for application and development purposes a double zeta (DZ) or triple zeta (TZ) type basis set is typically used. A polarized TZ basis set is often assumed to provide results close to the KS-limit, but no explicit calibration has been performed. Given that the basis set is an integral part of many developments of new exchange-correlation functionals, the total error becomes a combination of errors in the Hamiltonian and the basis set. If the functional contains empirical parameters, the fitting will to some extent compensate for inadequacies in the basis set, and the employed basis set thus becomes an integral part of the model.[11–13]

In recent work we have performed an analysis of the basis set convergence of the HF energy with a nuclear centered Gaussian type basis set.[18] Based on this analysis we proposed a new type of basis sets, denoted polarization consistent, which should provide a systematic convergence towards the basis set limit. In the present work we show that these basis sets after reoptimization of the exponents with a DFT method provides a hierarchy of basis sets for establish-

[a)] Electronic mail: frj@dou.dk

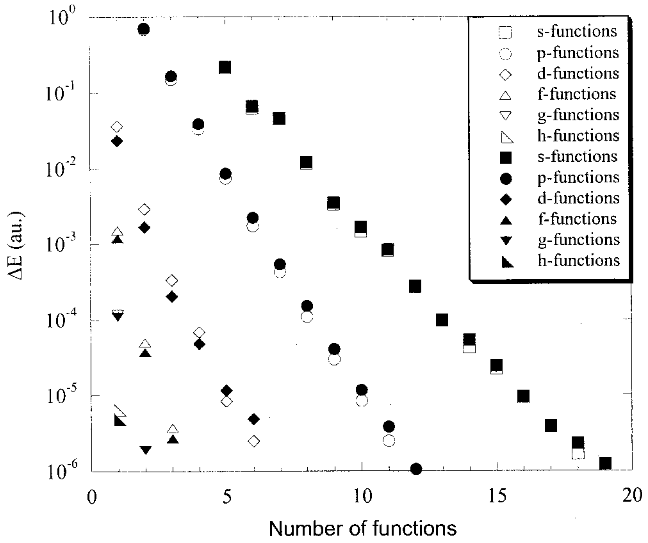

FIG. 1. Energy contributions for each basis function for the N_2 molecule. Open symbols are at the Hartree–Fock level, filled symbols are at the BLYP level.

ing the KS-limit, allowing a direct evaluation of the basis set error for other basis sets. In the present paper we focus on total atomization energies, comparisons for other properties will be reported in due course.

RESULTS

Defining polarization consistent basis sets

The principle for constructing the polarization consistent (pc) basis sets is that basis functions which provide similar amounts of energy should be included at the same stage, and each step up in quality adds a set of the next higher angular momentum functions. This is analogous to the procedure for constructing the correlation consistent basis sets, except that the analysis must be performed on molecular systems since the atomic energy is invariant to polarization functions. An analysis of a series of molecules at the HF level of theory showed that the optimum composition in term of functions with different angular momenta is insensitive to the molecular environment, and it is therefore possible to construct generally applicable atomic basis sets.[18] We thus proposed a series of polarization consistent basis sets pc-n ($n=0-4$), where the value n indicates the polarization level beyond the isolated atom. For first row elements a pc-0 basis set thus only contains s- and p-functions, a pc-1 basis set contains in addition d-functions, a pc-2 basis contains also f-functions, etc. The s- and p-exponents were optimized for the isolated atoms, while exponents for polarization functions were chosen based on explicit optimization for a series of representative molecules.

In general it is assumed that the basis set convergence for HF and DFT methods is very similar, since the three major components of the energy (electron kinetic energy, electron–nuclear attraction and electron–electron Coulomb energy) are identical in the two cases, and the similarity has been demonstrated explicitly for H_2.[19] In Fig. 1 we show a comparison of the energetic importance of each basis function for the N_2 molecule at the HF and BLYP (Becke gradient corrected exchange[6] and Lee–Yang–Parr gradient corrected correlation energy[7]) levels. All exponents have been explicitly optimized with the procedure used previously.[20] It is seen that the energetic importance is virtually the same for the two methods, and the pc-n basis set compositions in terms of the number of s-,p-,d-, etc. functions derived from HF results are thus also valid for DFT methods.

While the energetic importance of basis functions for the HF and DFT methods (illustrated by the BLYP results) are very similar, there are some minor differences in the values of the optimum exponents for the basis functions. The optimum exponents for the s- and p-functions (optimized for the isolated atoms) are in general slightly lower (more diffuse) at the BLYP level compared to HF, but there is only a very small dependence on the actual exchange-correlation functional used. The optimum polarization exponents were determined for a representative set of small molecules at the BLYP level, analogous to the previously used procedure.[18] The BLYP polarization exponents are in general slightly larger than the corresponding HF values, which to some extent is due to the more diffuse nature of the s- and p-functions. The final set of exponents for pc-n ($n=0-4$) basis sets for the elements H, C, N, O and F is given as supplementary material.[21]

Calibration

At the HF level of theory it is possible to establish the basis set limit for diatomic systems by performing numerical Hartree–Fock calculations.[22] Table I shows the errors in total energies calculated by the uncontracted pc-n basis sets, with the exponents taken as the BLYP optimized values. The atomic energies are calculated for spherical atoms, i.e., all p-orbitals are equivalent. It is seen that the error relative to the HF-limit decreases by roughly an order of magnitude for each step up in basis set quality, and extrapolation (discussed below) based on the pc-2, -3 and -4 data further improves the results by approximately a factor of 2. The extrapolated results have relative errors (absolute error divided by the total energy) of approximately 10^{-7}, which for these systems translate into absolute energy errors on the order of a few micro-hartree. The worst case system is H_2 with a relative error of 9×10^{-7}, primarily due to the choice of polarization exponents suitable for molecular calculations, which are not optimum for describing the short bond distance in H_2. The errors for the diatomic systems are approximately evenly distributed between the sp- and polarization spaces, as indicated by the atomic errors.

For application purposes the absolute energy is of little importance, since most properties of interest are related to energy differences. The total atomization energy, defined as the energy of the molecular system relative to the isolated atoms, will display a faster convergence with respect to the basis set size, as the molecular and atomic errors to some extent will cancel. Since the atomic energy only depends on the s- and p-functions, a large fraction of the remaining error will be related to the polarization functions. Alternatively, the total atomization energy can be defined relative to the corre-

TABLE I. Errors in Hartree–Fock total energies (atomic units) relative to the HF-limit with the pc-*n* basis sets. $\langle \Delta E_{rel} \rangle$ is the average relative error (absolute error divided by the total energy).

System	pc-0	pc-1	pc-2	pc-3	pc-4	xpol[a]	HF-limit[b]
H	0.003 397	0.000 927	0.000 089	0.000 003	0.000 000	0.000 000	−0.500 000
C	0.109 567	0.018 340	0.001 374	0.000 046	0.000 004	0.000 002	−37.688 619
N	0.172 944	0.029 271	0.002 133	0.000 067	0.000 005	0.000 002	−54.400 934
O	0.260 857	0.045 422	0.003 322	0.000 103	0.000 009	0.000 004	−74.809 398
F	0.373 936	0.066 484	0.004 771	0.000 143	0.000 013	0.000 007	−99.409 349
H_2	0.012 920	0.002 716	0.000 267	0.000 009	0.000 002	0.000 001	−1.133 630
C_2	0.246 076	0.041 619	0.003 274	0.000 126	0.000 013	0.000 007	−75.406 565
N_2	0.476 999	0.072 133	0.006 342	0.000 239	0.000 020	0.000 006	−108.993 826
O_2	0.595 291	0.104 083	0.008 911	0.000 332	0.000 029	0.000 011	−149.668 753
F_2	0.734 126	0.141 810	0.011 031	0.000 481	0.000 069	0.000 042	−198.773 443
NH	0.188 937	0.033 313	0.002 696	0.000 096	0.000 009	0.000 004	−54.978 585
CN	0.338 248	0.055 295	0.004 594	0.000 158	0.000 011	0.000 003	−92.225 134
FH	0.396 593	0.073 335	0.005 453	0.000 174	0.000 020	0.000 013	−100.070 802
CO	0.450 865	0.071 322	0.005 818	0.000 205	0.000 020	0.000 010	−112.790 907
NF	0.569 276	0.104 692	0.008 047	0.000 288	0.000 027	0.000 012	−153.842 418
$\langle \Delta E_{rel} \rangle$	4.4×10^{-3}	8.3×10^{-4}	7.0×10^{-5}	2.4×10^{-6}	2.6×10^{-7}	1.3×10^{-7}	

[a] pc-2, pc-3 and pc-4 results extrapolated with Eq. (4).
[b] References 23 and 24.

sponding diatomic molecules (H_2, C_2, N_2, etc.), which will allow some of the errors associated with polarization functions also to cancel. In general, the more similar the systems to be compared are, the greater the error cancellation can be expected. The energy difference between two different conformations of the same molecule, for example, is expected to show a fast basis set convergence. In this respect the total atomization energy defined relative to the isolated atoms can be considered as the most stringent test after total energies.

The error in total atomization energies (kJ/mol) at the HF level for the 10 diatomic systems in Table I are shown in Table II. The performance of each basis set is evaluated by calculating the mean absolute deviation (MAD) and maximum absolute deviation (MaxAD) relative to the numerical HF reference values. It is seen that each step up in basis set quality roughly increases the accuracy by an order of magnitude. With the pc-4 basis set the maximum error is 0.1 kJ/mol, and extrapolation (discussed below) further improves the results by approximately a factor of 2, giving an average error for the atomization energies of 0.01 kJ/mol. The error due to incomplete *sp*-function space is expected to show a good degree of cancellation for atomization energies, however, the error due to incomplete polarization space will not cancel. Since the latter is approximately half the error in absolute energies (*vide supra*), the error in atomization energy is expected to increase with system size, with a magnitude of approximately 0.01 kJ/mol or less per atom.

TABLE II. Errors in Hartree–Fock atomization energies (kJ/mol) relative to the HF-limit for the 10 diatomic systems in Table I. MAD = mean absolute deviation. MaxAD = maxium absolute deviation.

	Uncontracted		Contracted	
Basis	MAD	MaxAD	MAD	MaxAD
pc-0	116.03	344.23	121.76	361.23
pc-1	19.61	35.68	22.78	49.25
pc-2	2.85	5.95	2.03	5.98
pc-3	0.18	0.51	0.21	0.49
pc-4	0.03	0.11	0.04	0.11
xpol[a]	0.01	0.07	0.08	0.20
cc-pVDZ			26.56	37.37
cc-pVTZ			4.52	6.77
cc-pVQZ			1.03	1.83
cc-pV5Z			0.37	0.78
cc-pV6Z			0.10	0.24
xpol[b]			0.03	0.14

[a] pc-2, pc-3 and pc-4 results extrapolated with Eq. (4).
[b] cc-pVQZ, cc-pV5Z and cc-pV6Z results extrapolated with Eq. (1).

The contraction of the pc-*n* basis sets was discussed in the previous paper.[18] We proposed a general contraction scheme based on the expansion coefficients from calculations on the atomic systems. This has similarly been transferred to the current DFT basis sets by taking the coefficients from BLYP calculations on isolated atoms. The final contracted basis sets have been purified by the method of Davidson,[25] with functions having coefficients less than 10^{-5} being neglected, to provide the minimum number of primitive functions in each contraction. Contraction of the pc-*n* basis sets slightly degrades the performance, with the 3s2p1d contraction of the pc-1 basis set being the most problematic. Extrapolation based on the contracted results, however, actually degrades the performance over the raw pc-4 results. We also note that the computational saving by basis set contraction for HF and DFT calculations is not as large as for electron correlation methods, as the computational time for the former is dominated by integral evaluations. The comparison with the cc-pVXZ basis sets shows that the pc-3 basis set provides results intermediate between those from the cc-pV5Z and cc-pV6Z basis sets, while the pc-4 basis sets performs better than cc-pV6Z.

Estimating the Kohn–Sham limit

Since the current pc-*n* basis sets have been explicitly optimized for DFT methods, the above results at the HF level are likely to overestimate the error at DFT levels. We thus estimate that the pc-4 basis set in connection with extrapola-

TABLE III. Systems used for calibration.

Ionization Potential (42):

C, N, O, F, CH_4, NH_3, OH, H_2O, FH, C_2H_2, C_2H_4, CO, N_2, O_2, CO_2, CF_2, CH_2, CH_3, C_2H_5, CN,CHO, H_2COH, CH_3O, CH_3OH, CH_3F, CH_3CH_2OH, CH_3CHO, CH_3OF, NCCN, NH, NH_2, N_2H_2,N_2H_3, cyclopropene, allene, sec-C_3H_7, benzene, furan, pyrrole, toluene, phenol

Atomization Energies (95):

H_2, C_2, N_2, O_2, F_2, CH, CH_2(triplet), CH_2(singlet), CH_3, CH_4, C_2H, C_2H_3, C_2H_5, C_3H_5, NH, NH_2,NH_3, OH, H_2O, FH, C_2H_2, C_2H_4, C_2H_6, CN, NCCN, HCN, CO, HCO, H_2CO, CH_3O, CH_3OH,CH_3CO, H_2COH, N_2H_4, NO, NO_2, H_2O_2, CO_2, COF_2, N_2O, NF_3, O_3, F_2O, C_2F_4, CF_3CN, CH_2F_2,CHF_3, CH_3NH_2, CH_3CN, CH_3NO_2, CH_3ONO, CH_3CHO, CH_3CH_2O, CH_3CH_2OH, HCOOH,$HCOOCH_3$, CH_3CONH_2, propyne, allene, cyclopropene, cyclopropane, propane, butadiene, 2-butyne,methyl cyclopropane, bicyclobutane, cyclobutane, cyclobutane, isobutene, butane,isobutene, spiropentane, benzene, aziridine, dimethylamine, ethylamine, ketene, oxirane, glyoxal,dimethylether, vinylfluoride, acrylonitrile, acetone, acetic acid, acetyl fluoride, isopropanol, methylethyl ether, trimethylamine, furan, pyrrole, pyridine, sec-C_3H_7, tert-C_4H_9, toluene, phenol

Molecular Relative Energies (10):

CH_2(triplet)−CH_2(singlet), cyclopropene−allene, cyclobutene−bicyclobutane, cyclobutane−isobutene, butane−isobutene, CH_3NO_2−CH_3ONO, oxirane−CH_3CHO, CH_3CH_2OH−dimethylether, isopropanol−methyl ethyl ether, H_2COH−CH_3O

tion is capable of giving total energies accurate to 10^{-7} in a relative sense, and atomization energies accurate to better than 0.01 kJ/mol per atom. This is sufficiently accurate to provide a rigorous benchmarking of commonly used basis sets, and evaluate the performance of the pc-*n* basis sets and their contraction schemes. In order to provide a more representative sampling we have selected molecules containing the elements H, C, N, O and F from the G3 data set,[26] shown in Table III, providing a total of 42 ionization potentials and 95 atomization energies. Although electron affinities are also part of the G3 test set, this has not been considered at present. An accurate calculation of electron affinities is known to require diffuse functions, and such extensions will be considered separately. We have in addition also considered 10 relative molecular energies derived from the G3 data set for species with the same atomic composition. Such relative molecular energies are expected to converge faster with respect to the basis set size than atomization energies, as discussed above.

The geometry for all species have been taken as the B3LYP/6-31G(d,p) optimized. Open shell species have been treated within the UHF framework, including the isolated atoms, for which the wave function has inequivalent *p*-orbitals. All calculations have been performed with the GAUSSIAN 98 program package[27] with the default grid size for calculating the exchange-correlation term. We have tested that the results are stable toward the use of larger grids to within 0.04 kJ/mol in the worst case and better than 0.01 kJ/mol on average.

Table IV shows the composition and contraction for a selection of basis sets. The correlation consistent basis sets (cc-pVXZ) are designed for correlation energies, and are available up to $X=6$ for many elements.[3] The Pople style basis sets STO-3G,[28] 6-31G(d,p)[29] and 6-311G(2df,2pd)[30] basis sets are of minimum, double and triple zeta quality, respectively, and are very popular in routine applications. The 6-311+G(3df,2p) basis set has been used by Scuseria and co-workers for developing and testing new exchange-correlation potentials.[11] The corresponding Dunning–Huzinaga DZP and TZP basis sets[31] are also commonly used, and the TZ2P basis set has been used by Handy,[12] Tozer,[13]

TABLE IV. Basis set compositions.

	Contracted		Uncontracted	
Basis	M_A/M_H	Composition	M_A/M_H	Composition
pc-0	9/2	3s2p/2s	14/3	5s3p/3s
pc-1	14/5	3s2p1d/2s1p	24/7	7s4p1d/4s1p
pc-2	30/14	4s3p2d1f/3s2p1d	45/17	10s6p2d1f/6s2p1d
pc-3	64/34	6s5p4d2f1g/5s4p2d1f	84/38	14s9p4d2f1g/9s4p2d1f
pc-4	109/64	8s7p6d3f2g1h/ 7s6p3d2f1g	131/67	18s11p6d3f2g1h/ 11s6p3d2f1g
STO-3G	5/1	2s1p/1s	15/3	6s3p/3s
cc-pVDZ	14/5	3s2p1d/2s1p	26/7	9s4p1d/4s1p
6-31G(d,p)	15/5	4s2p1d/2s1p	28/7	11s4p1d/4s1p
DZP	14/5	3s2p1d/2s1p	29/7	9s5p1d/4s1p
GSAW-1	14/5	3s2p1d/2s1p	29/8	9s5p1d/5s1p
DFO-1	14/5	3s2p1d/2s1p	35-42/10	9-11s6-8p1d/4s2p
GSAW-2	18/6	4s3p1d/3s1p	33/8	10s6p1d/5s1p
TZ2P	27/9	5s4p2d/3s2p	38/11	10s6p2d/5s2p
cc-pVTZ	30/14	4s3p2d1f/3s2p1d	42/16	10s5p2d1f/5s2p1d
6-311G(2df,2pd)	30/14	4s3p2d1f/3s2p1d	43/16	11s5p2d1f/5s2p1d
DFO-2	28/12	4s3p3d/3s3p	56-63/18	12-13s8-10p4d/6s4p
6-311+G(3df,2p)	39/9	5s4p3d1f/3s2p	52/11	12s6p3d1f/5s2p
cc-pVQZ	55/30	5s4p3d2f1g/4s3p2d1f	68/42	12s6p3d2f1g/6s3p2d1f
cc-pV5Z	91/55	6s5p4d3f2g1h/ 5s4p3d2f1g	108/58	14s8p4d3f2g1h/ 8s4p3d2f1g
cc-pV6Z	140/91	7s6p5d4f3g2h1i/ 6s5p4d3f2g1h	161/95	16s10p5d4f3g2h1i/ 10s5p4d3f2g1h

TABLE V. Errors in BLYP ionization potentials (42 points), atomization energies (95 points) and relative molecular energies (10 points) (kJ/mol) for the systems in Table III relative to results obtained by extrapolation of pc-2, pc-3 and pc-4 energies. MAD = mean absolute deviation. MaxAD = maxium absolute deviation.

	Ionization potentials		Atomization energies		Molecular energies	
Basis[a]	MAD	MaxAD	MAD	MaxAD	MAD	MaxAD
pc-0	30.38	117.32	62.01	231.01	26.81	54.76
pc-1	6.79	26.14	18.37	42.19	4.13	7.57
pc-2	1.02	3.67	3.15	7.40	0.60	2.35
pc-3	0.11	0.53	0.12	0.72	0.07	0.26
pc-4	0.01	0.06	0.02	0.08	0.01	0.03
pc-0c	31.39	118.87	55.22	242.78	28.53	60.19
pc-1c	6.33	27.00	37.62	74.58	5.16	9.26
pc-2c	1.14	3.90	4.81	8.96	0.62	2.57
pc-3c	0.12	0.58	0.45	1.21	0.10	0.27
pc-4c	0.01	0.06	0.02	0.09	0.01	0.03
cc-pVDZ	25.45	65.25	40.87	82.23	6.24	11.55
cc-pVTZ	7.15	20.09	3.52	14.46	1.91	4.21
cc-pVQZ	2.93	8.44	2.88	10.86	0.75	1.59
STO-3G	178.89	452.32	376.54	1142.28	63.51	176.93
6-31G(d,p)	26.91	69.39	20.74	63.74	7.86	13.93
6-311G(2df,2pd)	11.79	32.22	11.31	48.57	3.00	7.02
6-311+G(3df,2p)	0.67	3.10	2.31	11.94	0.64	1.85
DZP	15.68	33.34	20.36	94.91	4.61	11.78
TZ2P	3.26	9.46	14.21	45.96	1.24	2.69
DFO1	4.61	14.35	22.41	47.69	2.53	5.18
DFO2	2.21	16.90	11.62	26.42	1.19	2.44
GSAW1	5.38	15.41	20.61	75.17	3.26	10.21
GSAW2	2.58	9.24	12.62	44.24	2.10	5.46

[a] pc-*n* denotes an uncontracted basis set, while pc-*nc* indicates the contracted version.

Thiel[15] and their co-workers for DFT development and testing purposes. Also included are some less common basis sets that have been proposed for DFT methods. The GSAW basis sets have been developed specifically for DFT calculations,[32] although they have not been widely used. More recently Porezas and Pederson have developed DFO basis sets by explicit exponent optimization at the DFT level.[33] The latter are somewhat different from commonly used basis sets, since the number of functions in each basis set depends on the element.

The MAD and MaxAD for the 42 ionization potentials, 95 atomization energies and 10 relative molecular energies, relative to results obtained by extrapolation of pc-2, pc-3 and pc-4 energies, are shown in Table V. The errors for the uncontracted pc-*n* basis set display the same convergence behavior as for the HF data in Table II, an error reduction by approximately an order of magnitude for each step up in basis set quality.

Contraction of a basis set is a compromise between computational efficiency and loss of accuracy. For the pc-0 and pc-1 basis sets a relatively large contraction error is acceptable, since the inherent error is fairly large, while only a small contraction error is consistent with the inherent high accuracy of the pc-3 and pc-4 basis sets. The suggested contractions of the pc-*n* basis sets based on a previous analysis[18] are shown in Table IV. Of these the contraction of the pc-1 basis to a DZP type (7s4p1d contracted to 3s2p1d) is the most problematic. This contraction increases the atomization energy by almost a factor of 2 relative to the uncontracted result, as seen in Table V. A contraction to 4s3p1d gives much better agreement with the uncontracted results (MAD = 19.18 and MaxAD = 44.27 kJ/mol). The results for the ionization potential and relative molecular energies do not show the large error increase by the 3s2p1d contraction. If the atomization energies are evaluated relative to diatomic reference data (H_2, C_2, N_2, etc.) instead, they do not show a similar large degradation by contraction. The 3s2p1d contraction error thus appears to be specific for atomization energies when atomic energies are used as the reference, which is the most stringent test, as discussed above. Since a 4s3p1d contraction increases the number of independent functions from 14 to 18 for each atom relative to a 3s2p1d contraction, we recommend the latter contraction, but users should be aware of the degraded performance for atomization energies.

The performance of other popular basis sets is also shown in Table V. The minimal STO-3G basis set performs much worse than the pc-0 basis set, although the latter has slightly fewer primitive basis functions. The main reason for the poor performance of the STO-3G basis set is the contraction to a minimal basis. Tests showed that a completely uncontracted version of the STO-3G basis set provides results comparable to those of the pc-0 basis set. We have previously shown that a contraction of the pc-0 basis set to a minimum 2s1p basis increases the error by roughly a factor of 3.[18]

The six basis sets of polarized double zeta quality, cc-pVDZ, DZP, GSAW-1, DFO-1, 6-31G(d,p), and pc-1, have comparable errors. The cc-pVDZ display the poorest performance, and the 6-31G(d,p) basis has significant errors for ionization potentials. Tests showed that the main reason for the poor performance of the cc-pVDZ basis is the contraction of the *sp*-functions. The pc-1 basis performs well, except for the above mentioned contraction problem for atomization energies. It should be noted that it performs better than the commonly used 6-31G(d,p) basis, despite the fact that pc-1 contains fewer functions.

Of the six polarized triple zeta type basis sets, cc-pVTZ, TZ2P, GSAW-2, DFO-2, 6-311G(2df,2pd) and pc-2, the pc-2 basis set in all cases provides significantly better results for a comparable number of basis functions (Table IV). The GSAW-2 and DFO-2 basis sets, which have been designed for use with DFT methods, are inferior to the pc-2 results. For the larger basis sets, the pc-3 basis set provides much better results than the cc-pVQZ, by approximately an order of magnitude, despite the comparable number of basis functions in the two basis sets. Table II indicates that the pc-4 basis set provides results better than those from the cc-pV6Z basis set.

The errors associated with the 6-311+G(3df,2p) and TZ2P basis sets are of particular interest since they have been used for developing and testing new exchange-correlation functionals. Scuseria and co-workers have shown that some of the most accurate functionals (e.g., VSXC, B3LYP and PBE1PBE) in connection with the 6-311+G(3df,2p) basis set give MAD values for atomization energies compared to experimental results of 10–20 kJ/mol, with corresponding MaxAD values of 30–40 kJ/mol.[34] These values can be compared with the MAD and MaxAD basis set errors of 2 and 12 kJ/mol (Table V). The MAD and MaxAD values for ionization potentials are 0.5–0.7 and 2–3 kJ/mol compared to experimental data, and Table V shows that the basis set alone provides a MAD of 0.7 kJ/mol and a MaxAD of 3.1 kJ/mol compared to the basis set limit. Handy and co-workers have used the TZ2P basis set for developing their parametrized HCTH functional, where 15 parameters are fitted to experimental data.[12] Their MAD for atomization energies in the final model is 24 kJ/mol, which can be compared with the MAD value of 14 kJ/mol due to basis set incompleteness (Table V). It would thus appear that a good part of the basis set error has been absorbed in the parametrization.

The present results suggest that the error from basis set incompleteness with commonly employed DZP or TZP type basis sets is not insignificant compared with the inherent error in some of the most accurate exchange-correlation functionals. This indicates that basis sets with smaller inherent errors should be used in future functional developments, as for example the pc-*n* basis sets. It should also be noted that for exchange-correlation functionals having empirical parameters, the basis set used in the parametrization becomes an integral part of the model, analogous to semi-empirical methods. When parameters are derived by fitting results from calculations with a specific basis set to experimental data, the parameters absorb some of the basis set error, and it is therefore possible that calculations with larger and more complete basis sets will actually degrade the performance.

Extrapolation procedures

Theoretical analysis suggests that the correlation energy calculated by wave mechanics converge as L^{-3}, where L is the highest angular momentum included in the basis set.[35] For the hydrogen atom an analysis by Klopper and Kutzelnigg suggest that the total (i.e., HF) energy has an exponential dependence on the square root of the number of *s*-type Gaussian functions.[36] No theoretical analysis is available for the dependence of HF or DFT energies on the highest angular momentum included in the basis set for molecular systems. Numerical results indicate that the L-convergence is also exponential, and a square root dependence appears to fit the data slightly better than a straight exponential.[20]

An exponential function of the type shown in Eq. (1) has been used in other applications for estimating the basis set limit for both HF, DFT and correlation energies using the cc-pVXZ basis sets.[4,17,37]

$$E=E_\infty+Ae^{-BL}. \tag{1}$$

In the present case extrapolation by Eq. (1) improves the cc-pVQZ results by approximately a factor of 2, except for relative molecular energies where the improvement is marginal.

Based on the numerical results for diatomic systems[20] we have also considered a corresponding extrapolation function depending on the square root of L:

$$E=E_\infty+Ae^{-B\sqrt{L}}. \tag{2}$$

From the principle of construction, we have argued that extrapolations of the type shown in Eqs. (3) and (4) should be suitable for extrapolating total energies from the pc-*n* basis sets (n_s is the number of *s*-functions in the basis set):[18]

$$E=E_\infty+A(L+1)e^{-Bn_s}, \tag{3}$$

$$E=E_\infty+A(L+1)e^{-B\sqrt{n_s}}. \tag{4}$$

Based on the theoretical analysis by Klopper and Kutzelnigg,[36] and the numerical data in Ref. 20, we prefer the function shown in Eq. (4) for extrapolation to the basis set limit. Furthermore, function (4) provides the best agreement with the numerical HF data in Tables I and II when extrapolating pc-2, -3 and -4 results, as well as pc-1, -2 and -3 results. When extrapolating the pc-2, -3 and -4 energies, however, there is little difference between the results from using either of the above four functions. The MAD from extrapolation by functions (1)–(3) differ by less than 0.01 kJ/mol from those obtained by function (4), and the corresponding MaxAD is less than 0.03 kJ/mol.

The performance of the extrapolation for the pc-0,-1,-2 and pc-1,-2,-3 total energies with the function in Eq. (4) is shown in Table VI. Results from pc-0,-1,-2 extrapolations show little or no improvement over the raw pc-2 results. The corresponding pc-1,-2,-3 extrapolated results represents an improvement over the raw pc-3 results, however, the improvement is not impressive, being less than a factor of 2. The performance of the extrapolation for the contracted pc-*n* basis set results follows those for the uncontracted results, except that the pc-2c,-3c,-4c extrapolated results provides little or no improvement relative to the raw pc-4c result. This is presumably due to the (relatively) large contraction

TABLE VI. Errors in BLYP ionization potentials (42 points), atomization energies (95 points) and relative molecular energies (10 points) (kJ/mol) for the systems in Table III, relative to results obtained by extrapolation of pc-2, pc-3 and pc-4 energies. MAD=mean absolute deviation. MaxAD=maxium absolute deviation.

	Ionization potentials		Atomization energies		Molecular energies	
Basis[a]	MAD	MaxAD	MAD	MaxAD	MAD	MaxAD
pc-1	6.79	26.14	18.37	42.19	4.13	7.57
pc-0,-1	5.43	16.93	26.88	82.02	3.32	10.40
pc-2	1.02	3.67	3.15	7.40	0.60	2.35
pc-0,-1,-2	0.78	3.63	2.43	9.10	0.91	2.46
pc-1,-2	0.68	2.61	1.88	6.08	0.48	1.91
pc-3	0.11	0.53	0.12	0.72	0.07	0.26
pc-1,-2,-3	0.08	0.37	0.28	0.65	0.03	0.07
pc-2,-3	0.08	0.42	0.18	0.49	0.05	0.14
pc-4	0.01	0.06	0.02	0.08	0.01	0.03
pc-3,-4	0.00	0.02	0.01	0.02	0.00	0.01
pc-1c	6.33	27.00	37.62	74.58	5.16	9.26
pc-0c,-1c	5.20	15.15	46.38	119.97	4.87	12.01
pc-2c	1.14	3.90	4.81	8.96	0.62	2.57
pc-0c,-1c,-2c	0.91	3.49	5.81	15.51	1.62	3.86
pc-1c,-2c	0.81	2.83	2.07	7.16	0.63	2.28
pc-3c	0.12	0.58	0.45	1.21	0.10	0.27
pc-1c,-2c,-3c	0.07	0.34	0.22	0.67	0.06	0.17
pc-2c,-3c	0.08	0.44	0.25	0.80	0.08	0.20
pc-4c	0.01	0.06	0.02	0.09	0.01	0.03
pc-2c,-3c,-4c	0.01	0.04	0.05	0.17	0.01	0.03
pc-3c,-4c	0.00	0.02	0.03	0.08	0.00	0.01

[a]pc-n denotes an uncontracted basis set, while pc-nc indicates the contracted version. pc-0,-1,-2 indicates a three-point extrapolation with Eq. (4) in the text. pc-0,-1 indicates a two-point extrapolation using Eq. (4) in the text with a B-parameter of 5.5.

error for the pc-2c basis, which destroys the fine balance required for attaining micro-hartree accuracies.

The main problem with the extrapolation functions of the type shown in Eqs. (1)–(4) is that they require three data points. The lowest order extrapolation is therefore based on pc-0, -1 and -2 results, however, the pc-0 results are so far from the limiting result that the extrapolated results do not improve raw pc-2 results. Extrapolation based on pc-1, -2 and -3 results gives a small improvement over the raw pc-3 results, but the relatively poor energies from the pc-1 basis set again prevent an efficient extrapolation.

The premise of extrapolation functions of the above type is that the B-parameter is relatively insensitive to the molecular system and the n-value in the pc-n basis set. Choosing the B-parameter to be a constant reduces the fitting function to a two-point extrapolation. Tests based on absolute energies compared to the numerical HF data in Table I, and by fitting to the BLYP energies obtained by a three-point extrapolation of pc-2, -3 and -4 results, suggest that the B-parameter to a good approximation can be taken as a constant with a value of 5.5. Results from such two-point extrapolations are also shown in Table VI. A two-point extrapolation of the pc-0 and pc-1 results actually deteriorates the performance for atomization energy, confirming the above conclusion that the pc-0 results are too far removed from the limiting value to provide a reliable extrapolation point. Extrapolation based on the pc-1 and pc-2 energies represents an improvement over the raw pc-2 results, and is also better than a three-point extrapolation of the pc-0, -1 and -2 data. Two-point extrapolations based on pc-2,-3 energies give results of similar quality as three-point extrapolation based on pc-1,-2 and -3 results. Since the overall computational cost will be dominated by the calculation with the largest basis set, only the pc-1,-2 two-point extrapolation is recommended for general use. If the pc-3 and pc-2 results are available, the corresponding pc-1 data can be generated by a marginal increase in computational cost, and used with a three-point extrapolation of the form in Eq. (4). Finally if pc-4 results are available, they can be used with the corresponding pc-2 and -3 data with a three-point extrapolation. Note that only the uncontracted versions of the basis sets should be used with the latter extrapolation. These recommended extrapolations in general improve the results by approximately a factor of 2.

The relatively small improvement by extrapolation is in sharp contrast to the situation for correlation energies, where extrapolation is an essential ingredient in obtaining high accuracy. The main difference between the two cases is the inherent fast basis set convergence of DFT (and HF) methods. Since each successive increase in quality of the pc-n basis sets gives approximately an order of magnitude improvement, even a two-point extrapolation procedure employs data which are at least a factor of 10 further removed from the limiting value. Given that the fundamental variable, the number of basis functions in each pc-n basis set, is quantized and only differ by 2–4 between the different n-values, this makes it difficult to design an extrapolation function capable of substantially improving the raw results.

CONCLUSIONS

It is shown that the previously proposed polarization consistent basis sets after reoptimization of the exponents and contraction coefficients at the DFT level provides a well-defined hierarchy for approaching the Hartree–Fock or Kohn–Sham basis set limit for molecular calculations. Each step up in basis set quality improves the results by approximately an order of magnitude. The largest pc-4 basis set provides results with an error in total atomization energy of less than 0.01 kJ/mol per atom. An exponential extrapolation can further improve the results by approximately a factor of 2. The pc-1 basis set is of polarized double zeta quality, and it is shown that it provides better results than other polarized double zeta type basis sets, for a smaller or comparable number of basis functions. The pc-2 basis set similarly provides results of higher accuracy than comparable sized basis sets of polarized triple zeta quality. The pc-3 and pc-4 results are superior to other standard basis sets for HF and DFT calculations. It is shown that commonly used basis sets for routine applications and for development purposes have basis set errors which are comparable to the inherent error in the most accurate exchange-correlation functionals.

ACKNOWLEDGMENTS

The author thanks J. Kobus for some of the numerical HF results in Table I, and C. Haettig for the atomic coefficients used in the contraction of the pc-*n* basis sets. This work was supported by grants from the Danish Natural Science Research Council.

[1] R. G. Parr and W. Yang, *Density Functional Theory* (Oxford University Press, Oxford, 1989).
[2] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic Structure Theory* (Wiley, New York 2000).
[3] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989); J. Phys. Chem. A **104**, 9062 (2000).
[4] 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).
[5] J. D. Perdew and Y. Wang, Phys. Rev. B **33**, 8800 (1986).
[6] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[7] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[8] T. Van Voorhis and G. E. Scuseria, J. Chem. Phys. **109**, 400 (1998).
[9] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996); **78**, 1396 (1997)(E).
[10] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[11] A. D. Rabuck and G. E. Scuseria, Theor. Chem. Acc. **104**, 439 (2000); Chem. Phys. Lett. **309**, 450 (1999); S. N. Maximoff and G. E. Scuseria, J. Chem. Phys. **114**, 10591 (2001).
[12] F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy, J. Chem. Phys. **109**, 6264 (1998); A. J. Cohen and N. C. Handy, Chem. Phys. Lett. **316**, 160 (2000); A. D. Boese, N. L. Doltsinis, N. C. Handy, and M. Sprik, J. Chem. Phys. **112**, 1670 (2000); A. D. Boese and N. C. Handy, *ibid.* **114**, 5497 (2001); W.-M. Hoe, A. J. Cohen, and N. C. Handy, Chem. Phys. Lett. **341**, 319 (2001).
[13] G. Menconi, P. J. Wilson, and D. J. Tozer, J. Chem. Phys. **114**, 3958 (2001); P. J. Wilson, T. J. Bradley, and D. J. Tozer, *ibid.* **115**, 9233 (2001).
[14] A. D. Becke, J. Chem. Phys. **107**, 8554 (1997); H. L. Schmider and A. D. Becke, *ibid.* **109**, 8188 (1998); A. D. Becke, J. Comput. Chem. **20**, 63 (1999).
[15] M. Filatov and W. Thiel, Phys. Rev. A **57**, 189 (1998).
[16] R. S. Fellers, D. Barsky, F. Gygi, and M. Colvin, Chem. Phys. Lett. **312**, 548 (1999); S. J. Jenkins and D. A. King, *ibid.* **317**, 381 (2000).
[17] K. S. Raymond and R. A. Wheeler, J. Comput. Chem. **20**, 207 (1999).
[18] F. Jensen, J. Chem. Phys. **115**, 9113 (2001).
[19] K. Aa. Christensen and F. Jensen, Chem. Phys. Lett. **317**, 400 (2000).
[20] F. Jensen, J. Chem. Phys. **110**, 6601 (1999); F. Jensen, Theor. Chem. Acc. **104**, 484 (2000).
[21] See EPAPS Document No. EJCP-SA6-116-305217 for supplementary tables. 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.
[22] J. Kobus, Adv. Quantum Chem. **28**, 1 (1997).
[23] J. Kobus (private communication).
[24] T. Koga, S. Watanabe, K. Kanayama, R. Yasuda, and A. J. Thakkar, J. Chem. Phys. **103**, 3000 (1995); J. Kobus, D. Moncreiff, and S. Wilson, Mol. Phys. **96**, 1559 (1999).
[25] E. R. Davidson, Chem. Phys. Lett. **260**, 514 (1996).
[26] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople, J. Chem. Phys. **109**, 7764 (1998); L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, *ibid.* **112**, 7374 (2000).
[27] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 98, Gaussian, Inc., Pittsburgh, PA, 1998.
[28] W. J. Hehre, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **51**, 2657 (1969).
[29] W. J. Ditchfield and J. A. Pople, J. Chem. Phys. **56**, 2257 (1971); P. C. Hariharan and J. A. Pople, Theor. Chim. Acta **28**, 213 (1973).
[30] M. J. Frisch, J. A. Pople, and J. S. Binkley, J. Chem. Phys. **80**, 3265 (1984); R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, Theor. Chem. Acc. **72**, 650 (1980).
[31] T. H. Dunning, J. Chem. Phys. **55**, 716 (1971); S. Huzinaga, *ibid.* **42**, 1293 (1965).
[32] N. Godbout, D. R. Salahub, J. Aldzelm, and E. Wimmer, Can. J. Chem. **70**, 560 (1992).
[33] D. Porezag and M. R. Pederson, Phys. Rev. A **60**, 2840 (1999).
[34] M. Ernzerhof and G. E. Scuseria, J. Chem. Phys. **110**, 5029 (1999); C. Adamo, M. Ernzerhof, and G. E. Scuseria, *ibid.* **112**, 2643 (2000).
[35] W. Kutzelnigg and J. D. Morgan III, J. Chem. Phys. **96**, 4484 (1992); **97**, 8821(E) (1992).
[36] W. Klopper and W. Kutzelnigg, J. Mol. Struct. **135**, 339 (1986); W. Kutzelnigg, Int. J. Quantum Chem. **51**, 447 (1994).
[37] A. Halkier, T. Helgaker, P. Jorgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998); 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, *ibid.* **111**, 9157 (1999); J. S. Lee and S. Y. Park, *ibid.* **112**, 10746 (2000); A. J. C. Varandas, *ibid.* **113**, 8880 (2000).