Basis sets for ADF and BAND

ADF Program System
Release 2009.01

Scientific Computing & Modelling NV
Vrije Universiteit, Theoretical Chemistry
De Boelelaan 1083; 1081 HV Amsterdam; The Netherlands
E-mail: support@scm.com

Copyright © 1993-2009: SCM / Vrije Universiteit, Theoretical Chemistry, Amsterdam, The Netherlands
All rights reserved

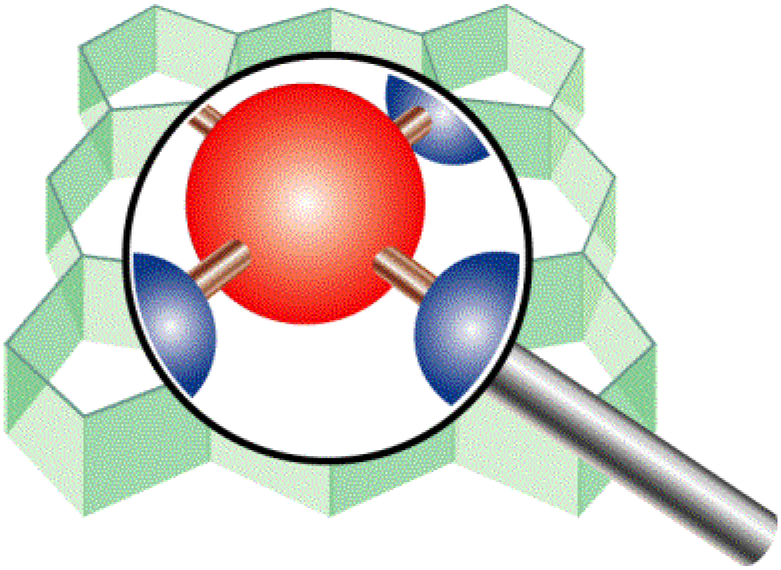

Table of Contents

Introduction

Purpose of this document

This document intends to do the following:

- Explain which basis sets are suitable for your problem.
- Describe the changes that have been made to the basis sets in ADF and explain why these changes have been made.
  Important: starting from the ADF2002 release the names of the standard basis sets have changed to more intuitive names:
  I→SZ, II→DZ, III→DZP, IV→TZP, and V→TZ2P.
  Similarly, the names for even-tempered basis sets in the $ADFRESOURCES/ET subdirectory have obtained similar names (VI→ET-QZ3P).
- Document the typical accuracy that can be obtained for diatomic molecules, using the various basis sets in ADF.
- Explain how even-tempered basis sets can be generated with the basis set utilities in ADF.

What should you certainly read?

If you are a new user of ADF, and no basis set expert, your primary interests should be in the advice what basis sets you should use and in studying some of the tables which summarize the quality of the new basis sets for a wide variety of diatomic molecules. This will give you an idea of what accuracy you can expect.

For people who have used ADF before, it is highly recommended to study which basis set files have been modified. In case you have performed calculations in which the basis set quality was very important it may be advisable to repeat some calculations with the improved basis sets in order to check that this does not influence your conclusions. This is only needed in case the basis sets you used have been substantially modified. Especially calculations on molecules containing P, Li, Be, K, or Ca, (or the heavier atoms Rb, Sr, Cs, Ba, Fr, Ra) atoms deserve careful attention.

Names of the standard basis sets have changed to more intuitive names: I→SZ, II→DZ, III→DZP, IV→TZP, V→TZ2P, and VI→ET-QZ3P.

Only for basis set experts or users with very specific demands on their basis sets, the description of the basis set utilities is of interest.

History, motivation, and general considerations

With the first release of ADF in 2002, the existing basis sets have been modified, but also new basis sets have been added. The reason for this is that we would like to facilitate calculations near the basis set limit, allowing users to remove this source of errors from their calculations (at the expense of CPU time). Two circumstances triggered our effort to test and improve the ADF basis sets. The first was the sabbatical of Prof. Del Chong (University of British Columbia, Vancouver, Canada) that he spent in the Theoretical Chemistry group in Amsterdam in the spring of 2000. Professor Chong, who has been involved in the development of basis sets for many years, pointed out to us the power and flexibility of even-tempered (ET) basis sets. These basis sets consist only of 1s, 2p, 3d, etc. functions, of which the factor between neighboring exponents for a certain l-value is a constant. Subsequently he started to develop such Slater Type Orbital (STO) basis sets for use in the ADF program. In the meantime the choice of polarization functions in ADF was reconsidered. The combination of these ET basis sets with improved polarization functions yielded good results in test calculations on molecular binding energies.

Another important factor was the ZORA quadruple zeta (QZ4P) basis developed by Dr. E. van Lenthe. These basis sets had been devised in a similar way as had been done earlier for basis sets I-V in ADF, by fitting each numerical orbital in an atom to several STOs. (Names of these standard basis sets I-V have changed to more intuitive names: I→SZ, II→DZ, III→DZP, IV→TZP, and V→TZ2P. Additionally a large number of polarization functions were added. These basis sets were intended for near basis set limit calculations on small systems. Subsequent tests were performed on many small molecules, in which the quality of various basis and fit sets was compared. The major overall conclusion was that, for normal properties like binding energies, both large ET basis sets and the QZ4P basis sets gave results very close to each other and very close to results obtained in the literature with very large GTO basis sets. This was a pleasing conclusion as we could then use these basis and fit sets to test the old ADF basis and fit sets. Many of such tests were performed by Dr. E. van Lenthe. In these tests occasional deficiencies in the old ADF basis sets were uncovered. Although no strict definition of deficiency was used, a basis set was considered suspect if it led to errors in the bond energy much larger than typical for that size of basis. In such a case the old, suspect basis set was adapted. Such an adaptation usually implied a slightly larger number of functions, or a more logical choice for the number of polarization functions or the exponent of the polarization functions. The changes that were made will be described more fully below.

Prof. Del Chong has devised two improved large basis sets for ADF that become available with ADF2002.03. The TZ2P+ basis set is available for the transition metals Sc-Zn only (Z=21-30). These basis sets are nearly identical to TZ2P except for a better description of the d-space (4 d-functions instead of 3). An improved even-tempered (ET) basis set has also been devised by Del Chong. It is named ET-pVQZ and available for the elements H-Kr. Because of smaller basis set superposition errors, the Et-pVQZ basis is to be preferred over the current ET-QZ3P and ET-QZ+5P basis sets. See also: D.P. Chong, E. Lenthe, S.J.A. van Gisbergen and E.J. Baerends, *Even-tempered Slater-Type orbitals revisited: From Hydrogen to Krypton.* Journal of Computational Chemistry **25**, 1030 (2004)

In the ADF2005.01 version the TZ2P+ basis set is now available also for the lanthanides (Z=57-70), although only in the ZORA directory. These basis sets are nearly identical to ZORA/TZ2P except for a better description of the f-space (4 f-functions instead of 3).

In the ADF2005.01 version augmented basis sets are available, for use in time-dependent density functional calculations. These are augmented SZ, DZ, DZP, TZP, TZ2P non-relativistic basis sets for H-Kr. They are named ASZ, ADZ, ADZP, ATZP, and ATZ2P, respectively. They are devised by Prof. Del Chong for use in TDDFT calculations, especially for the reasonable accurate calculation of excitation energies, with a relatively small basis set.

In the ADF2006.01 version non-relativistic all electron basis sets optimized for Hartree-Fock calculations are available in the subdirectory Special/Bunge. Ref. C.F.Bunge, J.A.Barrientos and A.V.Bunge, Atomic Data and Nuclear Data Tables 53,113-162(1993). Only to be used in atomic calculations, since polarization functions are missing.

Further details are given in the README files in the basis set directories.

New naming conventions and abbreviations

ADF uses Slater Type Orbitals (STOs). This is different from most other DFT programs, which usually employ Gaussian Type Orbitals (GTOs), for which many standard basis sets are available in the literature. The advantage of the STOs is that fewer of them are needed than with GTOs to get a reasonable description of the molecule. A disadvantage is that not much work has been done to develop balanced STO basis sets. We have attempted to remedy this with the presently available STO basis set database in ADF.

The ADF-BAND program uses Numerical Atomic Orbitals (NAOs) in addition to STOs in order to get an even better description with only a limited number of basis functions.

ET stands for even-tempered. This may apply to the basis and to the fit. In our ET basis sets, only 1s, 2p, 3d, and 4f functions occur. In the fit sets, 5g functions occur in addition to this. [Currently all ADF basis sets are restricted to f functions and fit sets are restricted to g functions as the highest l-value]. The exponents in ET basis sets are given by the simple formula:

$\zeta = \alpha * \beta^{l}$ where l=1,... N

Here ζ is the exponent of the STO, b (which should be larger than 1) defines how far apart two consecutive exponents are, and a determines what the most diffuse exponent is. In principle, each l-value has its own set of α, β, N.

Initially, the only basis sets provided with ADF were those in the directories I, II, III, IV, V, which now have the more intuitive names SZ, DZ, DZP, TZP, and TZ2P, respectively. The increasing numbers point to an increase in size and quality. It is not possible to give a formally correct short general classification for each basis set directory. However, generally speaking we can say that SZ (old name I) is a single-zeta basis set, DZ (old name II) is a double zeta basis set, DZP (old name III) is a double zeta polarized basis, TZP (old name IV) is a core double zeta, valence triple zeta, polarized basis set, and finally TZ2P (old name V) is a core double zeta, valence triple zeta, doubly polarized basis. This explains the more intuitive names that are given for the basis sets. The names have also been changed since some of the basis sets have been modified substantially.

The same is classification is used for the directories ZORA/SZ-TZ2P (old names ZORA/I-V). These were added later because of the special requirements on basis sets for ZORA relativistic calculations, especially in the core region. The ZORA/QZ4P basis set can be loosely described as core triple zeta, valence quadruple zeta, with four sets of polarization functions.

The TZ2P+ basis sets for the transition metals Sc-Zn are nearly identical to TZ2P except for a better description of the d-space (4 d-functions instead of 3). The ZORA/TZ2P+ basis sets for the lanthanides La-Yb are nearly identical to TZ2P except for a better description of the f-space (4 f-functions instead of 3).

AUG stands for augmented. These are augmented standard basis sets with some diffuse STOs are added, especially meant for TDDFT calculations. They are named ASZ, ADZ, ADZP, ATZP, and ATZ2P, which means augmented SZ, augmented DZ, augmented DZP, augmented TZP, and augmented TZ2P, respectively.

Credits

All basis sets in the ZORA directory have been built and tested by Erik van Lenthe, who also tested and improved many basis sets in the directories SZ-TZ2P (I-V) . Del Chong inspired the present efforts and provided his completeness profile utilities. Furthermore he devised and tested the ET basis sets on atomic energies and suggested how many polarization functions should be chosen for each element and how to choose their exponents properly. Stan van Gisbergen generated the basis set utility scripts and did some molecular tests on the ET basis sets. He also performed molecular tests to define new exponents for some ADF polarization functions. Pieter Vernooijs has performed much of the earlier work on basis and fit set generation in ADF. Evert-Jan Baerends provided much useful general advice on the proper construction of balanced basis sets.

References

A few key references that were very important in this project are:

The paper by Raffennetti on design and optimization of even-tempered STO basis sets[Raffenetti, 1973 #1143]. The paper by Del Chong describes completeness profiles as a visual tool in estimating the completeness of a basis set[Chong, 1995 #1215]. Finally, Zeiss and coworkers [Zeiss, (1979) #1216] describe field-induced polarization functions for STOs. These are useful for defining basis sets with diffuse functions for (hyper)polarizability and other property calculations.

More recent references:
E. van Lenthe and E.J. Baerends, *Optimized Slater-type basis sets for the elements 1-118.* Journal of Computational Chemistry **24**, 1142 (2003)
D.P. Chong, E. van Lenthe, S.J.A. van Gisbergen and E.J. Baerends, *Even-tempered Slater-Type orbitals revisited: From Hydrogen to Krypton.* Journal of Computational Chemistry **25**, 1030 (2004)

What basis SET SHOULD I USE IN ADF?

This question is hard to answer in general, but a few general suggestions can be made. This will be split here into several subsections.

ZORA or nonrelativistic calculation?

The first question to ask is: am I going to do a ZORA calculation to include scalar relativistic effects? This is recommendable for systems containing heavy nuclei, but there is no objection to doing a ZORA calculation for a system with light atoms only, as the required CPU time does not increase very much if ZORA is used.

If you are doing a ZORA calculation, you will need the ZORA basis sets which can be found in $ADFHOME/atomicdata/ZORA. You may also use the all electron basis sets from the ET or AUG directory, but be aware that these were optimized to non-relativistic calculations. Currently the ZORA basis sets cover the entire periodic table and besides all electron basis sets offer a choice of frozen cores. At present the ZORA directory does not contain basis sets with very diffuse functions, which may be required in calculations for hyperpolarizabilities or high-lying excitation energies, but for the lighter elements (H-Kr) you can certainly use the all-electron basis sets from the ET or AUG directory. Warning: in a ZORA calculation use only the frozen core basis sets coming from the $ADFHOME/atomicdata/ZORA directory, or use all electron basis sets.

If you do not use ZORA, your basis sets should come from the directories SZ, DZ, DZP, TZP, TZ2P (old names I-V), or one of the ET or AUG basis sets. Alternatively, you could make your own new ET basis set using the utility programs described below. For many of the heavy elements only ZORA basis sets are available, but for such elements it would be inadvisable to do nonrelativistic calculations anyway. For light elements the ZORA and normal basis sets should be identical except for the description of the frozen core. Usually the ZORA basis sets contain much steeper basis and fit functions to accurately describe the core region.

Large or small molecule?

In general it is advisable to use the best basis set that you can afford to use in terms of CPU time and memory. If you want to optimize the geometry or calculate the atomization energy of a diatomic molecule there is little reason not to use the very large ZORA/QZ4P basis, or (for light elements) a similarly large ET basis (we recommend the ET-pVQZ basis). If you are studying a molecule with 100 atoms or more, the use of such large basis sets does not only become prohibitive because of the required CPU time and memory, but it also is much less needed than for smaller systems. In medium-sized or large molecules even the moderately large basis sets will prove to be quite adequate because of the effect of basis set sharing. Each atom profits from the basis functions on its many neighbors. Additionally, if a large basis contains diffuse functions, linear dependency problems may occur. See also the input key DEPENDENCY in the ADF User Guide. In many cases basis DZ or DZP will give acceptable accuracy for calculations on large systems. If you are inexperienced it may be prudent to test a few different basis sets to get a feel for the size of basis set effects. To get a rough idea for the size of various basis sets, we mention here the number of functions for all-electron basis sets from the directories ZORA/SZ up to ZORA/QZ4P. For carbon, the number of functions is 5 (basis ZORA/SZ,I), 10 (DZ,II), 15 (DZP,III), 19 (TZP,IV), 26 (TZ2P,V), 43 (QZ4P). The same numbers for hydrogen are: 1 (SZ,I), 2 (DZ,II), 5 (DZP,III), 6 (TZP,IV), 11(TZ2P,V), 21(QZ4P). These numbers arise because ADF uses 'pure' d and f functions. In other words, 5 instead of 6 d functions are used and 7 instead of 10 f functions. Note that especially the jump from TZ2P (old name V) to QZ4P is quite steep.

In an overgeneralizing fashion we can state that the single zeta basis SZ (old name I) is hardly ever sufficient to get more than a qualitative picture and should be used only when larger basis sets are not affordable. The double zeta basis DZ (old name II) performs already much better and may give quite reasonable results, for example in geometry optimizations on large molecules. However, in more subtle situations, for example if hydrogen bonds are important, it is advisable to use at least one set of polarization functions. This is the double zeta polarized DZP basis set (old name III). Basis set TZP (old name IV) extends the valence space (but not the core space which remains double zeta) to triple zeta. In basis TZ2P (old name V) an additional polarization function is added. For example, hydrogen gets a d polarization function in addition to its p polarization function and carbon gets an f polarization function on top of a d polarization function. Several tests have shown that often the second polarization function is of more use when it has an l value one higher than the first polarization function. This is reflected in the choice just described. If another polarization function is needed it is usually best to add another one of the lowest l-value (2p+1d for hydrogen, 2d+1f for carbon). This choice has been made in the ET basis VI. There, sometimes even three d polarization functions were added, for example 3 p functions for Be, and 3 d functions for S. The reason for this is that S can occur in hypervalent species such as SF_6, which put special demands on the basis set. In the case of Be, the unoccupied p level is so close in energy to the occupied ones that it is sometimes called a valence level. Symantics aside, it is clear that a proper description of the p level of Be is very important and it is therefore not strange to add a third p function. In general, the unoccupied levels for the atoms on the left side of the periodic table are close to the occupied ones. This makes it necessary to add a few extra functions for the lowest unoccupied levels in order to get a description which corresponds to the general level of accuracy one expects from the hierarchical basis set names SZ-TZ2P (old names I-V). This fact was insufficiently recognized in the previous ADF basis sets. In our tests this led to higher errors in the energy for the diatomics containing for example Li, Be, K, Ca and so on. By adding appropriate functions we have corrected this. As a consequence, the basis set quality for a particular subdirectory is now more uniform throughout the periodic system. At the same time we have attempted to increase the number of functions in a systematic fashion so that each element is described by at least as many functions of a particular l value as its predecessor.

The ZORA/QZ4P basis sets might be roughly described as core triple zeta, valence quadruple zeta, with 4 polarization functions (2 d and 2 f functions for C, 2 p and 2 d for H). The fit sets corresponding to these basis sets are also much larger than the fit sets found in basis sets SZ-TZ2P (old names I-V). If one has doubts about the adequacy of a fit set for a certain element, this can be tested by replacing its fit set by the large one from the QZ4P directory. In the ZORA/QZ4P basis sets, the choice for the exponents of the polarization functions was done in a systematic, but somewhat hand-waving manner. For this reason the exponents were always rounded to half integers. Also the geometric mean of the exponents usually does not coincide with the choices made in directories SZ-TZ2P (old names I-V) and the ET basis sets. However, the fact that two polarization functions (with reasonable exponents) are present instead of a single one is far more important. A reasonable intermediate basis set, in size between V and QZ4P might be envisaged in which a single polarization function is added, as described above. This is roughly the choice for the polarization functions made in the ET directory VI.

Frozen core or all-electron?

In general we recommend the use of frozen core basis sets if available. Especially for the heavier atoms the number of functions is much smaller than for their all-electron counterparts. Our tests indicate that the error made by invoking the frozen core approximation is usually clearly smaller than the difference with respect to slightly higher quality basis sets. For the ZORA/QZ4P basis sets, only all-electron basis sets are available as these are intended for near basis set limit calculations only in which the CPU time is not a major concern.

Geometry optimizations involving atoms with a too large frozen core may give rise to numerical problems. In such cases it is recommendable to use a smaller frozen core. In previous occurrences we have removed such atomicdata files from the database. In the BAND basis sets such large frozen cores still exist in some cases. They should be handled with care. They have not been removed as they will only lead to problems in special cases and will reduce CPU times in many other calculations.

For accurate results on properties like nuclear magnetic dipole hyperfine interactions (ESR), nuclear quadrupole coupling constants, and chemical shifts (NMR), all electron basis sets are needed on the interesting atoms. For such properties tight functions might be necessary for high accuracy, especially in a ZORA calculation.

Diffuse functions needed?

For accurate results on properties like polarizabilities, hyperpolarizabilities, and high-lying excitation energies, diffuse functions are needed. This is especially true for calculations on small molecules. In larger molecules the nature of the relevant virtuals is much more 'molecular', much less Rydberg-like, so that the normal basis sets may be sufficient. Basically all properties calculated through the RESPONSE keyword may require diffuse functions. If you use the EXCITATIONS keyword, the necessity of diffuse functions depends on the type of excitation you are interested in. The lowest excitations do not require diffuse functions, but Rydberg excitations do.

In case of diffuse basis functions the risk of linear dependency in the basis increases. This can be checked, and corrected for with the DEPENDENCY keyword. It is recommended to use this keyword for all calculations involving diffuse functions. A good default setting is

DEPENDENCY bas=1d-4

However, it may be advisable to experiment a bit with the parameter, especially if many linear dependent combinations of AOs are removed. Using too many diffuse functions on a large molecule will lead to insurmountable numerical problems. In such a case it is not only useless, but even harmful, to add many diffuse functions.

In the previous release only some basis sets were provided which contained diffuse functions. These were gathered in the directory Vdiff. Now several ET basis sets have been developed for the elements up to Ar containing some or many diffuse functions. We recommend to use these instead of the Vdiff directory. Most of these basis sets are quite large and not very suitable for large molecules.

In ADF2005.01 augmented basis sets were added in the AUG directory, especially devised for use in in TDDFT calculations, such that one can do a reasonable accurate calculation of excitation energies, with a relatively small basis set, see D.P. Chong, *Augmenting basis set for time-dependent density functional theory calculation of excitation energies: Slater-type orbitals for hydrogen to krypton.* Molecular Physics **103**, 749 (2005)

Normal or even-tempered basis?

For normal calculations (these form the vast majority) we recommend the use of the optimized basis sets in the directories SZ-TZ2P (old names I-V) and, for ZORA calculations, ZORA/SZ-QZ4P (old names ZORA/I-V). These should be sufficient in accuracy for even very demanding users and are available for the entire periodic system (in the case of the ZORA basis sets). They are also available with a frozen core variety, which saves much CPU time.

The ET basis sets on the other hand are available only in all-electron form at the moment. Furthermore, most are pretty large (larger, but also better than TZ2P). Additionally, relatively large basis set superposition errors were detected for molecules containing atoms in the row K-Kr. For this reason we only recommend ET basis sets for the elements H-Ar at the moment. There they have yielded quite nice, near basis set limit, results for the G2 test set. For these light elements the ET basis sets can be comparable in quality to the ZORA/QZ4P basis, even though it is smaller. The ET basis sets are considered to be especially useful when diffuse functions are required. In that case it is very easy to adapt the original ET basis and fit set. The utilities provided for this in ADF will be described below in an Appendix. The ET basis set utilities will also prove useful for users who want to experiment with making their own basis sets, or who have very special demands on the basis and fit. The provided utilities automate much of the work needed to make new atomicdata files.

What Basis sets have changed in ADF2002?

Files added in Dirac, SZ-TZ2P (I-V), ZORA

Names of the standard basis sets have changed to more intuitive names: I→SZ, II→DZ, III→DZP, IV→TZP, and V→TZ2P. Several new atomicdata files were introduced in the Dirac, ZORA/SZ-TZ2P (old names ZORA/I-V), and SZ-TZ2P (old names I-V) directories. Many of these are related to new sizes of the frozen core, some elements were not previously available in that directory. This is especially true for the directory TZ2P (old nameV) which has been extended with the 3d transition metals.

New in Dirac: Ba.4p Cs.4p Fr.5p Ra.5p Rb.3p Sr.3p

New in (ZORA/)DZ (II): Ac.5d Ba.4p Cs.4p Fr.5p Ra.5p Rb.3p Sr.3p

New in (ZORA/)TZP (IV): Ac.5d Ba.4p Cs.4p Fr.5p Ra.5p Rb.3p Sr.3p

New in (ZORA/)TZ2P (V): Co Co.2p Co.3p Cr Cr.2p Cr.3p Cu Cu.2p Cu.3p Fe Fe.2p Fe.3p Mn Mn.2p Mn.3p Ni Ni.2p Ni.3p Ni_d9 Ni_d9.2p Ni_d9.3p Sc Sc.2p Sc.3p Ti Ti.2p Ti.3p V V.2p V.3p Zn Zn.2p Zn.3p

Most important changes in basis sets SZ-TZ2P (I-V)

The most important additional differences in the basis sets DZ-TZ2P are summarized here:

changed 3D polarization function for Na* and Mg*

changed value of exponent of 3D polarization function in P* (3D 1.35 → 3D 1.90)

changed 4F polarization function for Na*

changed 2S function for Na.1s (IV,V)

changed 2P functions for Li* (IV,V)

changed 2P functions for Be* (IV)

additional 2S function for F.1s (IV,V)

additional polarization functions for basis sets of: Rb* Sr* Cs* Ba* Fr* Ra*

Most important changes in basis sets ZORA/SZ-TZ2P (ZORA/I-V)

All-electron basis sets for light elements (H up to Kr) are identical for ZORA and nonrelativistic: ZORA/SZ,DZ,DZP,TZP,TZ2P = SZ,DZ,DZP,TZP,TZ2P (old names I-V)

The ZORA frozen core basis sets for H to Kr differ only in the core description from their analogues in SZ-TZ2P (old names I-V) The all-electron fit sets for Rb up to Uuo (DZ,TZP,TZ2P) have all been modified

Name of Y is Yttrium [was wrong in old basis set (Ytterbium)]

Most important changes Rb-Uuo:

Basis sets modified for Rb Sr Cs Ba Fr Ra (extra d and or f functions)

New basis sets have been added: Ba.4p Cs.4p Fr.5p Ra.5p Rb.3p Sr.3p (ZORA/DZ,ZORA/TZP,ZORA/TZ2P)

Frozen core actinides (ZORA/TZP) now has 3 6P functions instead of 2 6P functions

Pt now contains a 6P polarization function (ZORA/DZ,ZORA/TZP) (this was a bug in the old basis ZORA/II en ZORA/IV)

La now contains a deep 2P function (this was a bug in the old basis sets)

The elements Rf up to Uuo now contain deep 2P functions (ZORA/DZ,ZORA/TZP,ZORA/TZ2P) [this was a bug in old basis sets]

The elements In* up to Xe* now contain an additional 5D polarization function (ZORA/TZ2P)

What accuracy do the basis sets give?

Tests on many diatomics were performed by Dr. E. van Lenthe to test the various basis sets. We now document the results of some of these tests, in order to give a feeling for the quality that can be obtained from the various basis sets. See also: E. van Lenthe, E.J. Baerends, *Optimized Slater-type basis sets for the elements 1-118.* Journal of Computational Chemistry **24**, 1142 (2003)

Summary of test results

Tests for nonrelativistic calculations on 36 diatomics containing oxygen, namely the oxides of the first 36 elements (H-Kr). All-electron basis sets were used. The new ZORA/QZ4P basis set was used to define the basis set limit result. Note that after these tests the fit sets in the ZORA/QZ4P basis set were slightly modified. The numbers in the table refer to bonding energies in eV. Differences were taken between the QZ4P results and the results in smaller basis sets. By construction, the errors in the QZ4P column are zero. Names of the standard basis sets have changed to more intuitive names: I→SZ, II→DZ, III→DZP, IV→TZP, V→TZ2P, and VI→ET-QZ3P.

	QZ4P	II	DZ	III	DZP	IV	TZP	V	TZ2P	VI
	new	old	new	old	new	old	new	old	new	new
Average error	0.0	1.47	1.33	0.48	0.39	0.27	0.18	0.19	0.06	0.01
Average absolute error	0.0	1.47	1.33	0.48	0.39	0.27	0.18	0.19	0.06	0.02
Maximum error	0.0	4.53	2.84	1.72	1.07	1.31	0.41	1.21	0.13	0.18
Worst case	all	Cao	SO	CaO	BeO	CaO	FO	CaO	O_2	CaO

A few comments are in order to explain this table.

> The oxides were used as a small test set because their equilibrium bond lengths are known in many cases. Also, they have a large influence on the electronic structure of the molecule, so that they also test the adequacy of the polarization functions.
>
> The errors in the small basis sets are systematic, because the isolated atoms are described reasonably well, but the molecular energy is not deep enough. For this reason the average errors and average absolute errors are (nearly) always equal.
>
> In the old basis sets (and also in the new ET basis set VI) CaO is the worst case. This has been solved in the new basis sets by adding 3d functions. The average errors with respect to basis QZ4P go down from 1.47 eV (II) to 1.33 eV for basis DZ, from 0.48 eV (III) to 0.39 eV for basis DZP, from 0.27 eV (IV) to 0.18 eV for basis TZP and from 0.19 eV (V) to 0.06 eV in basis TZ2P.
>
> The improvement in the average errors has been achieved by dealing with the worst cases. For this reason the maximum error increases even more significantly than the average error. For example the largest error in basis V has gone down from 1.21 to 0.13 eV in basis TZ2P.
>
> The ET basis VI has a much lower deviation from QZ4P than TZ2P. This is probably mainly due to more polarization functions. The small difference between VI and QZ4P indicates the reliability of both basis sets.
>
> The results for frozen core basis sets I-V in comparison with the new SZ-TZ2P basis sets (not shown) are very similar to those shown here for all-electron basis sets and can therefore be warmly recommended.

Test calculations on 100 diatomics containing oxygen, using all-electron ZORA basis sets. Many basis sets for (very) heavy elements are included here, which could not be included in the table above. The numbers have the same interpretation as above and are again in eV.

	QZ4P	DZ	DZ	TZP	TZP	TZ2P	TZ2P
	ae	fc	ae	fc	ae	fc	ae
Average error	0.00	0.95	1.07	0.20	0.20	0.05	0.05
Average absolute error	0.00	0.98	1.07	0.20	0.21	0.05	0.05
Maximum error	0.00	2.86	2.83	0.74	0.74	0.19	0.17
Worst case	all	SO	SO	UuoO	UuoO	ThO	UuoO

Again we place a few comments on these frozen core and all-electron results.

The trends are very similar to those in the previous table for the lighter elements.

The frozen core results are very satisfactory, as they are very close to the results with the corresponding all-electron basis sets. The error introduced by the frozen core approximation is typically much smaller than the one introduced by basis set incompleteness.

The average errors are quite comparable to those from the previous table. The heavier elements do not seem to be much more difficult than the lighter ones.

For heavy elements no reliable ET basis set is yet available for comparison.

More results, all-electron, nonrelativistic on roughly 140 different diatomics at experimental or 'reasonable' equilibrium geometries.

	QZ4P	II	DZ	IV	TZP	VI
	new	old	new	old	new	new
Average error	0.00	0.95	0.89	0.17	0.11	0.00
Average absolute error	0.00	0.95	0.89	0.17	0.11	0.02
Maximum error	0.00	4.53	2.84	1.31	0.32	0.18
Worst case	all	CaO	SO	CaO	O_2	CaO

Only the nonrelativistic basis sets DZ (old name II) and TZP (old name IV) are fairly complete for heavier elements.

In the ET basis VInew, the Ca polarization functions have not yet been extended. For that reason the difference with respect to QZ4P is nonzero.

Also for these general diatomics (not just oxides) the average and maximum errors have decreased substantially, especially for basis TZP.

Same table, but now for frozen core basis sets. In all these tests the smallest frozen core files were employed (i.e. the largest basis).

	QZ4P	II	DZ	IV	TZP
	new	old	new	old	new
Average error	0.00	1.13	0.73	0.55	0.13

Average absolute error	0.00	1.14	0.75	0.59	0.16
Maximum error	0.00	9.43	2.87	9.43	1.80
Worst case	all	BaO	SO	BaO	ThO

The frozen core approximation has little influence on the accuracy for the new basis DZ, but a somewhat larger effect on the new basis TZP. This is especially due to certain worst cases, such as ThO.

ZORA, all electron, over 240 diatomics

	QZ4P	II	DZ	IV	TZP	V	TZ2P
	new	old	new	old	new	old	new
Average error	0.00	0.83	0.70	0.23	0.11	0.12	0.02
Average absolute error	0.00	0.83	0.70	0.23	0.11	0.12	0.03
Maximum error	0.00	6.39	2.83	3.36	0.44	2.02	-0.16
Worst case	all	Bao	SO	BaO	I_2	BaO	Cr_2

The average error goes down very nicely from 0.70 to 0.11 to 0.03 eV when going from DZ to TZP to TZ2P. The average error in basis TZ2P is clearly below 1kcal/mol (the famous chemical accuracy). Errors due to deficiencies in current xc functionals are still much larger than this. As a consequence, the ZORA/TZ2P basis will be more than adequate for all standard calculations.

It is to be expected that these conclusions will not dramatically change if larger test molecules are used. Also for geometry optimizations the improved basis sets SZ-TZ2P and ZORA/SZ-TZ2P should be more than sufficient for all standard cases. The ZORA/QZ4P can be considered a very safe (though expensive) option for basis set limit calculations.

Appendix. How TO make EVEN-tempered basis/fit sets?

The programs described below have not yet been made generally available.

The standard basis sets that are provided are sufficiently flexible to accommodate the needs of almost every type of calculation of almost every user. Therefore, the first question should be: do I really need to make my own basis set? Almost always, the answer should be 'no', because of the availability of the directories (ZORA/) SZ-QZ4P (old names I-V) and basis set directories containing diffuse ET basis and fit sets. If however, you decide that these basis sets might be insufficiently reliable for your purpose, you can use the utilities described below.

At the moment, there are several restrictions to these utilities. (We currently do not yet make these utilities generally available as they have not been very extensively tested yet. People who think they may need these utilities should contact SCM). First, only starting points (basis sets for the occupied shells) are available for elements up to Kr. Also, we currently have some reservations about the ET basis sets from K-Kr, because of the relatively large basis set superposition errors that occur. Nevertheless, for accurate all-electron nonrelativistic basis sets up to Kr, the utilities may still prove a very useful tool. The tools have not yet been tested to work properly for the generation of basis sets suitable for ZORA calculations.

What is the basic idea behind the basis set utilities?

The basis set utilities generate ADF atomicdata files for ET basis and fit sets, using some simple input from the user. With the first utility, the user selects the ET basis set for the occupied shells that should form the starting point. These were developed and tested on atomic total and orbital energies by Prof. Del Chong during a sabbatical spent in Amsterdam. These basis sets are intended to be at least of the quality of basis V, but usually better. However, also some smaller ET basis sets were made for more economical calculations. After selecting the basis set for the occupied shell, the user has to specify whether additional polarization, diffuse, or contracted basis functions should be added (also tight functions for ZORA calculations can be added, but this has not been properly tested). To answer these questions requires of course some expertise in basis sets. It is therefore recommended for users with some experience in this area. However, suitable, safe defaults are defined and suggested in the scripts. The user also defines the quality of the fit set which is desired. Only the highest quality fit set has been thoroughly tested. The other options are not supported at the moment. We will now describe the input and output of the various utilities in some detail.

The utility questbas

This utility is available in $ADFBIN/questbas.exe. Running this executable results in the following questions:

Please provide the atomic number of the element

For Carbon, we type 6

As a next question a long list of default basis sets is presented. These correspond to predefined choices for the number of diffuse, tight, and polarization functions, and are used to automatically generate all kinds of ET basis sets. If any of these standard basis sets are chosen, the answers to all questions are filled in automatically.

Choose 0 for the general basis set procedure

SUGGESTED CHOICE: 0

Here, we follow the suggested choice which allows us to explain all options, and type 0.

Now the script tells us we have the choice between 4 different basis sets for the occupied shells. The basis sets which have been most thoroughly investigated at present are the ET basis sets in which the factor beta (β) [defining the spacing between two subsequent STOs] is fixed at the value 1.7 for all l-values in the basis. These basis sets form the starting point for the directories VI and VII. For carbon, the output tells us that we would start from a 5s4p basis with this choice. In another ET basis set, with a different design philosophy, β is variable (typically 1.55 for s, 1.7 for p, 2.0 for d), but α is fixed. These basis sets occur in a large (valence quadruple zeta, core double zeta) and medium (augmented double zeta) variety. In the case of carbon, they are of the size 6s4p and 4s3p. These variable β basis sets have been less intensively tested in molecular calculations than the fixed β variety. For that reason a final verdict has to be postponed as to which type is preferable. The fixed β seems the safest choice at present, because it has been used more in molecular calculations. The large variable β basis leads to names like VIB and VIIB. We type here 1 for the fixed β variety. As output we now get information on our intermediate basis set

As a starting point you have selected an even-tempered basis with following values for n, alpha, beta:

S: 6 0.6592211993160408 1.700000047683716

P: 4 0.5316027776326796 1.700000047683716

The script follows with the following question:

You can specify here how many additional fus you want for each l-value. Suggested: 0 0 0 0

This option allows the user to add functions to the final basis by keeping the endpoints, most diffuse and tighest functions, fixed and reducing the β value. This option is not often useful and we ignore it by typing 0 0 0 0

The next question concerns the ZORA option. As this has not yet been thoroughly tested we choose 0 for a nonrelativistic basis set.

The next question is if we want to add functions. This is definitely needed, because the basis without polarization functions is quite poor, so we type 1.

Then we are asked how many polarization functions should be added. As a default is suggested 0 2 1, meaning 2d and 1f polarization functions. This means that we get more polarization functions than in basis V if we use the default. The geometric mean of the exponent of the polarization functions is stored inside the program in a data statement. This gives the 'best' exponent for a single polarization function. If more than polarization function is used, this will be the middle of these polarization functions. The beta value for the polarization function is taken identical to the beta value of the highest occupied l-value.

After typing the suggested default of 0 2 1, we are asked if we want to add diffuse functions. Whether such functions are needed in your application is discussed elsewhere in this document. Let us add one diffuse p function by typing 0 1 0 0

We are then asked about additional tight functions to improving the description of the core region. In this example, we do not add such functions by typing 0 0 0 0

Then we are asked what size of fit set we require. We type 3 to get the largest fit set, the only thoroughly tested option.

Finally we are asked if we want to let the program add diffuse function to make the basis set suitable for (hyper)polarizability calculations. We follow the suggested choice by typing 0 0. Otherwise, the utility would add diffuse functions until the most diffuse function is more diffuse than a stored default diffuse value (provided by Prof. Del Chong, based on field-induced polarization functions).

The output of the questbas utility is given in the (local) file raf_in. This file gives the input for the rafbas utility described next.

The utility rafbas

The utility $ADFBIN/rafbas.exe reads its input from the local file 'raf_in' and writes its output to the local file 'et_in'. No further user input is required. This utility basically transforms the answers of the user to the questions posed in 'questbas' into a set of N, α, β values for both the basis and the fit.

The rafbas and the other utilities were originally also intended for generating so-called cusp-satisfying basis sets, which consist of (for the s functions) 1 1s functions to which ET 3s functions are added. However, some tests indicated that the core description of these basis sets, which was supposed to be very good, did not improve upon those of the normal ET basis sets and therefore the cusp-satisfying basis sets were abandoned.

The rafbas utility contains information about the beta values for the fit. For the default 'VERYLARGEFIT' option, the beta values were chosen such that the overlap between two successive fit functions does not exceed 0.95 for s and 0.90 for higher l it functions. Earlier experience with generating fit sets led to the conclusion that even larger fit sets have a large risk of becoming linearly dependent which causes numerical problems. The rafbas utility also adds the polarization function exponents. It contains a list of 'best' exponents for many elements. It further more handles the options to extend the basis by decreasing beta and to add first- and second-order FIPs (field-induced polarization functions) for the generation of basis sets for (hyper)polarizability calculations.

The utility etprog

Similarly to the 'rafbas' utility, 'etprog' reads from the local file 'et_in' and writes to the local file 'et_tmp'. It also prints some output which gives information on what the meaning is of the numbers in the file 'et_in'. The result in 'et_tmp' is almost of the form (N-values basis, α-values basis, β-values basis, N-values fit, α-values fit, β-values fit).

The main task of etprog is to generate a fit set corresponding to the basis set information generated by rafbas. To this purpose, it first generates the most diffuse and most contracted products of basis functions. This defines the range which should at least be well described by the fit set. However, the range is extended somewhat more to further increase the quality of the fit. Experience shows that, at least for light elements, this indeed leads to very small fit errors in molecular calculations, at the expense of a large fit. The beta values for the fit are read from the input file and originate from a data statement in the rafbas program.

The utility etwrite

No calculations are needed in the utility 'etwrite', it merely writes out the final information in a usable form to several files. This utility reads from the local file 'et_tmp' and writes to the file 'C' because this was the element we selected. This file 'C' is an atomicdata file as required for an ADF create run and can be used without further change. This file also contains information about the final basis and fit set, which can be used to uniquely specify this basis and fit in a publication. The output further consists of local files like 'SSTO.BAS.INP' which is intended as input for Chong's completeness profile utility $ADFBIN/ssto.exe. The result of that program is a datafile which can them be visualized with the gnuplot viewing program using the other output file 'gnu.in.basis' as an example. For the interpretation of the completeness profiles, we refer to Prof. Chong's publications on this subject.

An example of how all these scripts and programs can be combined is to be found in $ADFHOME/ examples/ basis_fit_utils/e_make_bas_Kr/run, which should be more or less self-explanatory using the information given here. The directory $ADFHOME/examples/basis_fit_utils also contains other scripts which can be of use in the generation and testing of basis and fit sets.