

# Error-Balanced Segmented Contracted Basis Sets of Double- $\zeta$ to Quadruple- $\zeta$ Valence Quality for the Lanthanides

Rebekka Gulde,<sup>†</sup> Patrik Pollak,<sup>†</sup> and Florian Weigend<sup>\*,†,‡</sup>

<sup>†</sup>Institut für Physikalische Chemie, Abteilung für Theoretische Chemie, Karlsruher Institut für Technologie, Kaiserstraße 12, 76131 Karlsruhe, Germany

<sup>‡</sup>Institut für Nanotechnologie, Karlsruher Institut für Technologie, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

## S Supporting Information

**ABSTRACT:** For lanthanides, segmented contracted Gaussian basis sets of double- $\zeta$  valence to quadruple- $\zeta$  valence quality are presented, with two different polarization sets for each level of quality. The bases are designed for use in connection with small-core Wood–Boring effective core potentials. A set of compounds representing most lanthanides in their common oxidation states is used to assess the quality. Parameters investigated were atomization energies, dipole moments, and structure parameters for Hartree–Fock, density functional, and correlated (Møller–Plesset) methods. So, the “def2” basis set series is extended to lanthanides with errors that are very similar to those previously obtained for the other elements with this type of basis set. Furthermore, for lanthanides, auxiliary bases for density fitting of Coulomb and Hartree–Fock exchange matrices are presented and tested.

## I. INTRODUCTION

In a preceding work,<sup>1</sup> a system of error-balanced segmented contracted basis sets of double- $\zeta$  to quadruple- $\zeta$  valence quality was provided for the entire periodic table, except for lanthanides and actinides. Meanwhile, this basis set system became a quite frequently used tool, for which the extension to f elements appears reasonable. It bases on former developments of polarized double- $\zeta$  “def-SV(P)”, polarized triple- $\zeta$  “def-TZVP”, and polarized quadruple- $\zeta$  “def-QZVP” valence bases<sup>2</sup> (“def” is an abbreviation for “default”). An improvement of these sets regarding the error balance across the periodic table was obtained by using a large set of molecules that represents each element in (nearly) all of its common oxidation states. At each level of quality, two (sometimes identical) sets of polarization functions were provided: a larger one for post-Hartree–Fock (post-HF) calculations and a smaller one for treatments at the level of density functional theory (DFT). At the DFT level, def2 bases show typical deviations to the basis set limit, in the following termed “typical errors”, of ca. 20/5/1 kJ/(mol atom) in atomization energies with lower polarized def2-SV(P)/TZVP/QZVP bases. The term “typical error” was introduced previously<sup>1</sup> for brief characterization of the basis set quality; it is the sum of the absolute mean value ( $|\mu|$ ) of errors and its standard deviation ( $\sigma$ ). At the level of Møller–Plesset perturbation theory (MP2), the higher polarized def2-SVP/TZVPP/QZVPP bases must be used. Errors then are ca. three times larger than those for DFT. At the HF level, errors are similar to those for DFT, if the higher polarized sets are used; for the lower polarized sets, they are somewhat larger. It was demonstrated<sup>1</sup> that the error balance is reached to a large extent, i.e., that those values are largely independent of the type of elements (main group, d-metal, s-metal) present in the compound. For atoms beyond Kr, the def2 bases were designed for use in combination with scalar-relativistic effective core

potentials (ECPs). The ECPs are of the Wood–Boring (WB) type,<sup>3</sup> in case of the s- and d-elements, and of Dirac–Hartree–Fock (DHF) type<sup>4</sup> for the p-elements. Meanwhile, also DHF ECPs<sup>5</sup> and the respective segmented contracted error-balanced bases<sup>6</sup> are available for s- and d-elements. It was demonstrated<sup>6</sup> that the differences (in ground-state properties) between these two types of ECPs are rather small, compared to, for example, the errors of DFT or MP2.

For lanthanides, the number of available basis sets is much smaller. Examples are comparably large all-electron bases of atomic natural orbital type by Roos et al. (“ANO-RCC”),<sup>7</sup> the bases by the Sapporo group,<sup>8</sup> “SARC-ZORA” sets by Neese et al.,<sup>9</sup> as well as the Douglas–Kroll–Hess-adapted sets by Dolg et al.,<sup>10</sup> which are of identical size as the sets in ref 9, but show smaller errors. In particular, for heavy elements such as lanthanides, the employment of ECPs is a great practical advantage. Examples for large-core ECPs with shells up to 5s/5p/4d modeled in the potential are ECPs by Ross et al.,<sup>11</sup> and for ECPs also covering the f electrons, that by Dolg et al.<sup>12</sup> Small-core WB-ECPs covering the inner 28 electrons were developed by Dolg, Stoll, and Preuss<sup>13</sup> ~20 years ago; respective well-performing generally contracted<sup>14</sup> and segmented contracted<sup>15</sup> bases of roughly triple- $\zeta$  quality were provided more recently by Cao and Dolg (CD).

Still missing for the treatment of lanthanides with these well-established pseudo-potentials is a system of basis sets with different sizes and accuracies, like the def2 series, or the system of correlation consistent generally contracted bases started by Dunning<sup>16</sup> and continued by Peterson.<sup>5</sup> Nevertheless, it is important to balance the quality of basis sets over the compound to be treated, as the following example shows.

Received: April 12, 2012

Published: August 8, 2012



The bond energy of NaCl ( $d_{\text{Na-Cl}} = 237.72$  pm) at the DFT level (BP86 functional<sup>17</sup>) amounts to 196.8 kJ/(mol atom), if def2-QZVPP bases are employed; this needs 128 basis functions. If one uses def2-TZVPP bases, 74 basis functions are needed, and the error (difference to def2-QZVPP) amounts to  $-1.2$  kJ/(mol atom). Combining def2-QZVPP for Na and def2-SV(P) for Cl requires approximately the same number of basis functions, 76, but leads to a significantly larger error:  $+6.1$  kJ/(mol atom).

In the present work, we extended the system of “def2” type bases by designing error-balanced, segmented contracted basis sets in double- $\zeta$  to quadruple- $\zeta$  quality for lanthanides. These bases are developed for the small-core WB-ECPs of ref 13. These ECPs were chosen in order to be consistent with ECPs for the other elements. The respective bases by Cao and Dolg (CD)<sup>15</sup> served as an excellent starting point. We further present density fitting (or “resolution of the identity”, RI) bases for the Coulomb and exchange part of the Fock matrix.

In section II, we start with some general considerations about polarized  $n$ -tuple zeta valence bases, discuss the test set and reference bases, and describe the optimization procedure. In section III, accuracy for the newly developed as well as for the CD bases is discussed. In section IV, auxiliary basis sets for Coulomb and exchange operators are presented and their performance is assessed.

## II. DESIGN AND OPTIMIZATION

**A. General Considerations.** It is a commonly accepted definition of a Gaussian “ $n$ -tuple zeta valence basis” to include (i) one (contracted) Gaussian type orbital, cGTO, for each core shell; (ii)  $n$  GTOs (or cGTOs) for each valence shell; and (iii),  $(n - 1)$ ,  $(n - 2)$ , ... polarization functions with angular momentum quantum numbers exceeding  $l_{\text{max},v}$  by 1, 2, ..., with  $l_{\text{max},v}$  being the largest angular momentum quantum number of the valence shells. Using step (i), the chemically less active inner shells are described economically; step (ii) ensures that the flexibility for the description of chemically active valence shells is adjustable; and step (iii) allows correlation-consistent polarization to be realized, if the exponents were optimized by maximization of the correlation energy.

Unfortunately, for many different reasons, many deviations from this definition occur, at least beyond the first period (i.e., beyond Ne). Some examples follow. For alkaline and alkaline-earth metals,  $l_{\text{max},v} = 0$ , but, as the subsequent p-, d-, or f-shells are energetically close to the s-shell, they are often partially occupied in compounds and thus the number of polarization functions must be much higher than that according to step (iii). In addition, optimization should be done by minimizing the energy of the respective excited state, rather than by maximizing the absolute correlation energy. In case of the second period, it is important to provide polarizing d-functions also for the inner p-shell; otherwise one gets significant errors already at the HF level (e.g., some kJ/mol per atom in atomization energies, which are not tolerable for triple- $\zeta$  valence and larger bases).<sup>18</sup> Similar is true for the polarization of the 3d shells of 4p elements. On the other hand, for d-elements, usually less polarization functions than those due to step (iii) are needed.<sup>1</sup> Strictly following steps (i) and (ii) is possible to a large extent in the case of generally contracted basis sets, where one common set of exponents is used for all shells of the same angular quantum number (with different contraction coefficients for each shell), but not for segmented contracted sets. For the latter, orbitals are fitted by segments,

which leads to multiple use of cGTOs as the following example will show. The 1s shell shows a radial Slater decay; the first cGTO is first used to model this shell. The 2s shell has one radial node. The inner part of the 2s shell (between the origin and just before the first radial node) also shows Slater decay, so the first cGTO is additionally used to model this part of the 2s shell; similar correspondence holds for the outer part of the 2s shell and the middle part of the 3s shell, etc. This concept leads to comparatively short contraction lengths, which is advantageous for efficiency in many quantum chemical program packages. However, the multiple employment of cGTOs for similar but not identical shapes requires additional functions in particular for heavier elements and if accurate total energies are desired; bond energies are less critical, as these errors mainly concern inner shells (and the inner part of the atomic valence shell), and thus (largely) cancel out for bond energies. Furthermore, since cGTOs are multiply used, contraction coefficients and exponents must be optimized simultaneously, in contrast to generally contracted sets.

Thus, in the case of segmented sets for heavier elements, it is reasonable to associate labels such as “double- $\zeta$  valence”, etc. with specific basis set errors rather than with specific contraction schemes. Of course, it is also reasonable to keep the basis set size for the desired error at a minimum and use the same or at least similar contraction schemes for a group of elements (in the present case, for the lanthanides).

The employment of ECPs causes further complications. The inner shells are missing in the explicit description and the node structure of the outer shells is nonphysical. This, of course, makes the above-mentioned multiple employment of segments in its strict form impossible. Nevertheless, the similarity of consecutive shells of a specific angular quantum number (e.g., 4s/5s/6s) is great enough to at least partly allow the multiple use of segments.

**B. Molecular Test Set and Reference Bases.** Basis set quality is crucially determined by the ability to describe different electronic situations of the respective element with approximately the same quality. Hence, it is reasonable to define a set of test compounds that represent possibly each element in its common oxidation state and additionally to vary the type of bond partners. For instance, the electron shift from the metal to the bond partner is much larger for F than that for Br, despite the fact that, for both, the formal oxidation state of the metal atom in the respective compound is the same. Furthermore, also some lanthanide dimers were included, because they turned out to be particularly problematic. The resulting set covered the following 49 molecules: CeF, CeH<sub>2</sub>, CeO, CeF<sub>3</sub>, PrCl, PrH<sub>2</sub>, PrF<sub>3</sub>, NdCl, NdH<sub>2</sub>, NdO, NdF<sub>3</sub>, PmF, PmO, PmH<sub>2</sub>, SmF, SmH<sub>2</sub>, SmCl<sub>3</sub>, Eu<sub>2</sub>, EuCl, EuF<sub>2</sub>, EuH<sub>2</sub>, Gd<sub>2</sub>, GdF, GdF<sub>2</sub>, GdH<sub>2</sub>, GdF<sub>3</sub>, TbF, DyF, DyF<sub>2</sub>, DyCl<sub>3</sub>, HoO, HoF<sub>3</sub>, ErCl<sub>2</sub>, ErF<sub>3</sub>, TmCl, TmF<sub>2</sub>, Yb<sub>2</sub>, YbCl, YbH<sub>2</sub>, Lu<sub>2</sub>, LuF, Lu<sub>2</sub>O, Lu<sub>2</sub>N, LuF<sub>2</sub>, LuH<sub>3</sub>, LuF<sub>3</sub>, LuCl<sub>3</sub>, LuBr<sub>3</sub>, LuI<sub>3</sub>. Orbital occupations were determined at the DFT(BP86) level using CD bases for the lanthanide atoms and def2-QZVP bases for the others with the “fractional occupation number” approach;<sup>19</sup> the pseudo-temperature was reduced during iterations, so that integer occupations fulfilling the *aufbau* principle were reached in all cases. Afterward, structure optimizations were carried out at this level. The resulting structure parameters are listed in the Supporting Information, as well as the electronic configuration reflected by the number of unpaired electrons of the molecule as well as by orbital occupations of the lanthanide atom obtained at the HF level with reference bases (see below). Note

**Table 1.** Contraction Patterns for the Basis Sets def2-SV(P) to def2-QZVPP and Auxiliary Basis Sets RI-J and RI-JK Presented Here, as Well as That for the Bases from ref 15<sup>a</sup>

basis	spdf	ghi		<i>nb</i> f(AO/CAO) <sup>b</sup>	
		P	PP	P	PP
SV(P)/P	(14s 12p 9d 7f)/[10s 6p 4d 3f] {5111111111/531111/6111/511}		(1g)/[1g] {1}	69/82	78/97
TZVP/PP	(14s 14p 10d 8f)/[10s 7p 5d 4f] {5111111111/6311111/61111/5111}	(1g)/[1g] {1}	(3g)/[2g] {21}	93/116	101/131
QZVP/PP	(16s 18p 11d 10f)/[11s 8p 6d 5f] {6111111111/93111111/611111/61111}	(3g)/[2g] {21}	(6g 1h)/[3g 1h] {411/1}	118/151	138/187
CD <sup>c</sup>	(14s 13p 10d 8f)/[10s 8p 5d 4f] {5111111111/61111111/61111/5111}	(6g)/[3g] {411}		114/149	
reference	(31s 25p 18d 16f)/[31s 25p 18d 16f]	(6g 2h 1i)/[6g 2h 1i]		397/534	
RI-J	(18s 7p 6d 6f)/[14s 7p 6d 6f]	(6g 3h 5i)/[5g 2h 2i]		200/304	
RI-JK	(20s 15p 14d 13f)/[15s 13p 14d 13f]	(9g 8h 6i)/[13f 9g 7h 5i]		438/690	

<sup>a</sup>The prefix “def2” was omitted for brevity. In some cases (typically Er–Lu) for the def2-orbital bases, one p set was either added to the basis or released from the contraction. <sup>b</sup>The term *nb*f denotes the number of basis functions in the basis of spherical harmonics (AO) or Cartesian functions (CAO). <sup>c</sup>Cao and Dolg.<sup>15</sup>

that, from this point in the paper, the prefix “def2” is omitted for the sake of brevity.

The reference bases are of an even-tempered type and are constructed similar to that previously described.<sup>6</sup> The factor between subsequent exponents was set to  $10^{-1/4}$ , the smallest exponent was chosen to be smaller than the smallest of the respective CD basis by  $10^{-1/4}$ , the largest larger than the largest of the CD. For p-functions, two smaller exponents were added, as this turned out to significantly improve results. Furthermore, two h-functions and one i-function were added, leading to typically 31s 25p 18d 16f 6g 2h 1i reference basis sets. Finally, the exponents were scaled by a factor specific for each angular momentum type (and each element), which was determined by minimization of the atomic HF ground-state energy. For the remaining atoms, QZVPP bases were taken as reference. With these bases and the respective ECPs,<sup>13</sup> HF, DFT, and MP2 calculations were carried out for the test set with fixed structure parameters, yielding reference bond energies and dipole moments. For selected (worst) cases, reference structure parameters also were determined afterward. At the DFT level, the functional BP86<sup>17</sup> was used throughout; at the MP2 level, the frozen core approximation was applied for lanthanides to shells up to 4spd, for N, O, F to the 1s shell, and for Cl/Br/I to shells up to 2/3/4p. For the sake of clarity, we note that MP2 calculations for a specific basis are always carried out for the HF reference obtained with the same basis.

**C. Optimization and Specification of Basis Sets.** As a first step, the quality of the CD bases was tested, as those bases are quite proven in praxis and thus are very promising candidates for starting points. Contraction patterns of these basis sets, as well as of those developed in this work, are listed in Table 1. Bond energies per atom and dipole moments were calculated (with QZVP bases for the nonlanthanides) and statistically evaluated.

For bond energies, CD bases show errors, with respect to the reference basis, of  $0.64 \pm 1.41$  kJ/(mol atom) at the DFT level,  $-1.40 \pm 1.93$  kJ/(mol atom) at the HF level, and  $0.78 \pm 3.30$  kJ/(mol atom) at the MP2 level (i.e., “typical errors” of ca. 2/3/4 kJ/mol per atom, which is slightly more than triple- $\zeta$  quality, even better at the MP2 level); for individual results, see Tables S3–S5 in the Supporting Information. Concerning dipole moments, performance is of double- $\zeta$  valence (DFT:  $0.232 \pm 0.191$  D) or less (HF:  $0.561 \pm 0.533$  D); for individual

results, see Tables S7 and S8 in the Supporting Information. The reason for this turned out to be a lack of diffuse p-functions; we note, in passing, that this problem in the meantime was also recognized by Dolg.<sup>10</sup> Moreover, it became evident that reduction of the comparably large 4g set to 2g (MP2) or 1g (DFT) does not significantly reduce the quality in praxis.

Thus, the TZVP and TZVPP sets were derived from the CD bases by adding diffuse p sets, by slightly changing the contraction pattern of the p-set from {61111111} to {6311111} for Ce–Dy and to {63111111} for Er–Lu, and by reducing the polarizing g-set from {411} to {1} for TZVP and to {21} for TZVPP in the following manner: for the 1g set of TZVP, the geometric mean value of the two more diffuse g-functions of CD was taken. The additional 2g set of TZVPP was derived from the steeper 4g set of CD by forming the mean values of the first two and the second two primitives with the respective contraction coefficients as weights.

Exponents and contraction coefficients of the (14s 14p 10d 8f)/[10s 7p 5d 4f] set (plus 1p for Er–Lu) were reoptimized at an unrestricted HF level for the respective HF ground state (usually  $s^2d^{1p^{n-1}}$ , except for Nd, Sm, Eu, Yb). The resulting sets are of very similar quality as CD concerning bond energies, but in the typical TZVP/PP range for dipole moments (see below).

The double- $\zeta$  valence sets SVP and SV(P) were derived from the triple- $\zeta$  valence bases by removing one cGTO from the p/d/f-set of TZVP and keeping the s-set unchanged. The parameters of the resulting (14s 12p 9d 7f)/[10s 6p 4d 3f] set (plus 1p for Er–Lu) were fully optimized for the respective HF ground state. For the SVP set, the same 1g set as that for the TZVP set was used; for the lower polarized SV(P) set, it was omitted.

For QZVP/PP sets, not only energy differences but also total (HF) energies are desired to be close to the limit. Errors for first- and second-row elements are much smaller than 1 mE<sub>H</sub> (1 mE<sub>H</sub> = 27.211 meV = 2.6255 kJ/mol); for heavier elements, errors of  $\sim 1$ –2 mE<sub>H</sub> are achieved. This increase is not surprising, because complexity increases with the number of electrons, in particular, if ECPs are employed (because, in this case, the nonphysical inner part must be described). Furthermore, f-shells additionally increase complexity. Thus, it was decided to ease criteria and allow for errors of ca. 3–4 mE<sub>H</sub>. For comparison, we note that the CD sets show errors of

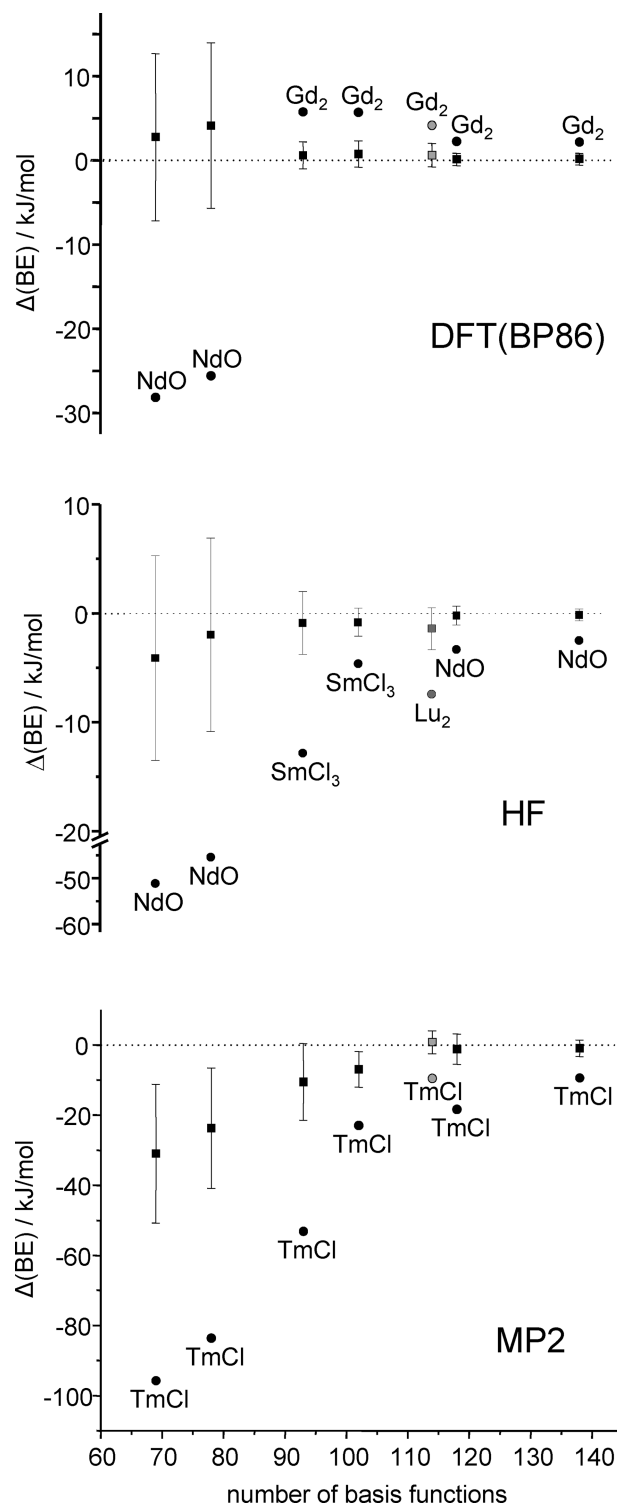


up to 16 (Dy)  $mE_H$ . In order to fulfill these requirements, the number of primitives had to be increased, compared to TZVP, by 2/4/1/2 for s/p/d/f-sets. The number of cGTOs was regularly increased by one for each angular quantum number. Polarization sets of QZVP are identical to that of TZVPP; for QZVPP, the original set from CD was taken plus one 1 h-function, which was roughly optimized by maximization of the MP2 correlation energy (because the higher polarized sets are intended mainly for correlated treatments). The resulting basis set parameters are sampled in the Supporting Information (Table S12), as well as all single data used for the statistical evaluations discussed in the next section (see Tables S1–S11 in the Supporting Information).

### III. RESULTS

**A. Atomic Energies.** Average errors in total unrestricted Hartree–Fock (UHF) energies for QZVPP/TZVPP/SVP bases amount to  $2.6 \pm 0.9/8.4 \pm 3.5/33 \pm 18 mE_H$ , worst cases are Ho ( $3.7 mE_H$ )/Dy ( $15 mE_H$ )/Yb ( $69 mE_H$ ) for the respective bases. At DFT, level errors are larger, because the bases were not explicitly optimized for this purpose. Average errors amount to  $13 \pm 9/32 \pm 27/65 \pm 40 mE_H$ , with worst cases Lu ( $29 mE_H$ )/Tb ( $86 mE_H$ )/Yb ( $123 mE_H$ ). Results for CD are very similar to that of TZVPP: at the HF level, the average error is  $8.8 \pm 4.2 mE_H$ , the largest 16  $mE_H$  (Dy), at the DFT level, the respective data are  $32 \pm 26 mE_H$  and 84  $mE_H$  (Tb). These errors are not unexpected, as the number of electrons to be described is quite large (30–43). For comparison, we note that, for Br (35 electrons), the errors for QZVPP/TZVPP/SVP bases amount to 0.6/50/366  $mE_H$  at the HF level and to 10/66/379  $mE_H$  at the DFT level, if a fully decontracted cc-pVSZ basis<sup>20</sup> is used as reference.

**B. Bond Energies.** Bond energies—in contrast to total energies—are measurable quantities and, thus, of much larger practical relevance. In Figure 1, we display, for HF/DFT/MP2 bond energies per atom, mean errors and standard deviations from the mean values over the test set, together with the “worst cases”. Data are additionally listed in Table S6 in the Supporting Information. At all levels of computation, one observes a reduction of typical errors when switching from a smaller basis to a larger basis within the def2 series developed here. The only exception is the change from SV(P) (without polarizing g) to SVP (with one polarizing g) at the DFT level, which does not improve results. Furthermore, the typical errors obtained with the def2 bases for the lanthanides match well with those obtained previously for the s/p/d-elements with def2 bases. The CD sets, which are of slightly smaller size than the QZVP sets, are of similar accuracy as the latter, in the case of MP2, but display TZVP to TZVPP quality for DFT and HF. Further trends are similar to those observed previously for the s/p/d-elements. Errors (absolute of the mean value plus the standard deviation) of the SV(P) basis amount to 10–20 kJ/mol at the DFT and HF levels and, thus, are smaller than typical errors of the method; for MP2, this goal is reached at triple- $\zeta$  valence quality. With triple- $\zeta$  valence bases (HF, DFT) and quadruple- $\zeta$  valence bases (MP2), errors on the order of 2–5 kJ/mol are reached, which is sufficient for most practical applications. Furthermore, at the DFT level, the larger polarization set does not provide a significant advantage over the smaller set. However, at the MP2 level, as well as at the HF level, the larger set improves results. For HF and DFT, increasing the basis mainly reduces the standard deviation, since, at the double- $\zeta$  valence level, mean errors already are



**Figure 1.** Deviations to the reference in bond energies per atom ( $\Delta(BE)$ , in kJ/mol) at the DFT (BP86) level, the HF level, and the MP2 level obtained with basis types SV(P), SVP, TZVP, TZVPP, QZVP, and QZVPP (from left to right, in black; the prefix “def2-” has been omitted) versus the number of spherical harmonic basis functions. In addition, data for the bases from ref 15 (Cao and Dolg, CD) are shown (in gray). The mean value over the test set is represented by a square, and the standard deviation from this value is represented by the error bar. The compounds showing the largest  $\Delta(BE)$ , “worst cases”, are listed explicitly (filled circles).

small; in contrast, for MP2, mean errors also systematically become smaller for larger bases.

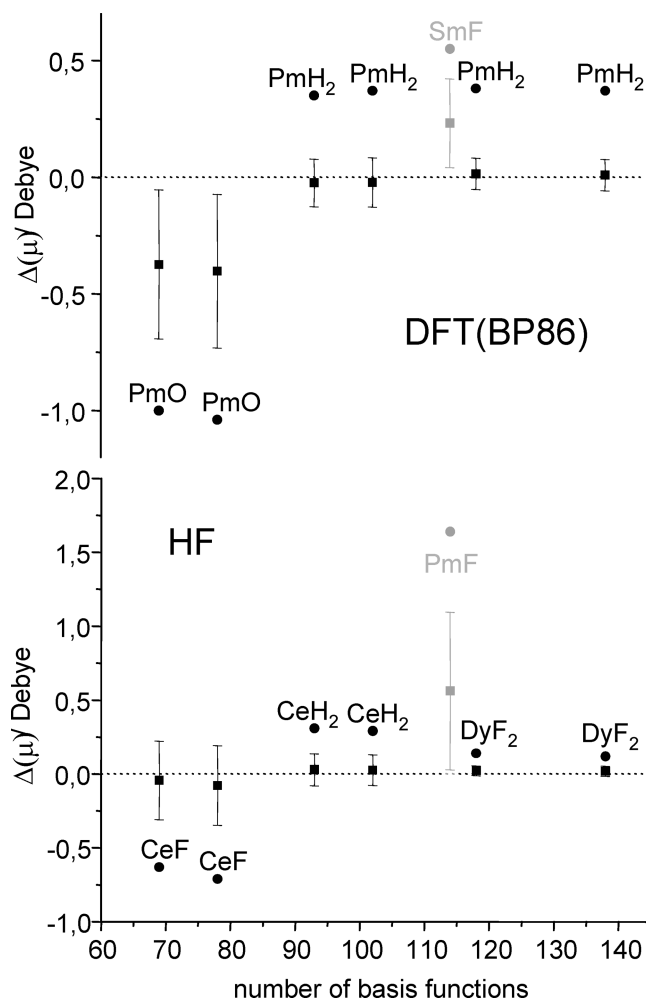
An important advantage of small-core ECPs compared to large-core ECPs is the possibility to cover different occupations of the *f*-shells. This, of course, requires a certain amount of flexibility of the basis set; otherwise, the description becomes biased toward the cases with occupations same as for the atoms. We thus have a detailed look at the bond energies of PrCl, ErCl<sub>2</sub>, PmF, TmCl, and TmF<sub>2</sub> (cf. Table S2 in the Supporting Information), which have an additional *f*-electron compared to the atoms (also see Tables S3–S5 in the Supporting Information). At HF and DFT, these compounds behave well, and all errors are as small as requested above; the worst case for DFT is ErCl<sub>2</sub> (errors of 0.35/1.1/5.4 kJ/(mol atom) at QZ/TZ/SV level), and the worst case for HF is TmCl (1.1/2.8/5.0 kJ/(mol atom)). MP2 results reveal a different picture: errors for PmF and PrCl amount to ca. 3/10/40 kJ/(mol atom) and thus are still tolerable, but for TmCl, we get 9.4/23/84 kJ/(mol atom), which, for the QZVPP basis particularly, is more than desired. We note in passing that this molecule is also problematic with CD bases, showing an error of ca. 9.6 kJ/(mol atom), which is similar to that which applies for ErCl<sub>2</sub> and TmF<sub>2</sub>. These problems, which obviously mainly occur at the MP2 level for the heavier lanthanides, can be solved by partial decontraction of *f*/*g*-shells to 511111/2211. The errors at the MP2 level for these modified QZVPP sets amount to 3.6/2.5/3.1 kJ/(mol atom) for ErCl<sub>2</sub>/TmCl/TmF<sub>2</sub>. We finally decided to maintain the contraction scheme for consistency and costs; nevertheless, in the case of changing *f*-occupations in heavy lanthanides, the above partial decontraction is advisable.

**C. Dipole Moments and Equilibrium Distances.** We continue the assessment of quality by considering first-order derivatives of the total energy: errors of dipole moments and equilibrium distances are discussed now. For the dipole moments, statistical values refer to the subset of 41 compounds that have a nonvanishing dipole moment. For HF and DFT, these data are displayed in Table 2 and are plotted in Figure 2

**Table 2. Statistical Data (Mean Values ( $\mu$ ) and Standard Deviations ( $\sigma$ )) for Errors in Equilibrium Distances (pm) with Respect to the Reference Basis at the DFT (BP86) Level, the HF Level, and the MP2 Level for All Diatomic Compounds That Are Listed as “Worst Cases” in Figures 1 and 2, Except for Gd<sub>2</sub> and Lu<sub>2</sub>**

	DFT		HF		MP2	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
QZVPP	−0.13	0.07	−0.02	0.04	0.21	0.08
QZVP	−0.08	0.10	0.03	0.12	0.44	0.35
TZVPP	−0.36	0.20	−0.23	0.27	0.09	0.51
TZVP	−0.33	0.21	−0.21	0.30	0.20	0.57
SVP	−2.06	1.11	−1.16	0.96	−1.27	1.06
SV(P)	−1.55	1.03	−0.4	0.93	0.65	0.80
CD	−0.22	0.13	0.13	0.13	0.33	0.18

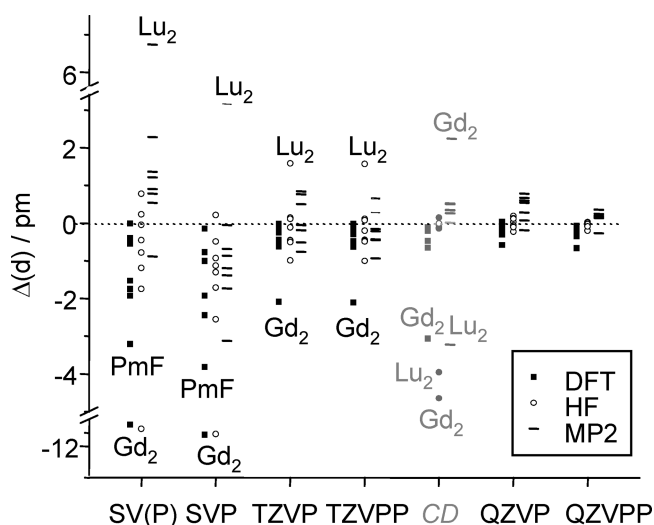
(additionally listed in Table S9 in the Supporting Information). Quadruple- $\zeta$  (and also triple- $\zeta$ ) valence bases yield very small typical errors of below 0.08 (0.15) D for both HF and DFT. For HF also at the double- $\zeta$  valence level, errors are very reasonable, ca. 0.3 D. This error almost completely arises from the standard deviation; thus, almost no systematic errors occur. At the DFT level, in contrast, a systematic underestimation by ca. 0.4 D is observed, which, together with a standard deviation



**Figure 2.** Deviations to the reference in dipole moments ( $\Delta\mu$ , in Debye) for the test compounds with nonzero dipole moment at the DFT (BP86) and HF level. (Also see Figure 1.)

of 0.3 D (similar to that of HF), amounts to comparably large typical errors of ca. 0.7 D. The reason for this is found in the basis set for the nonlanthanide atoms: If one changes the basis for O in PmO (the worst case) from SV(P) to TZVP and keeps the SV(P) basis for Pm, the error is reduced from 1.00 D to 0.46 D. A similar shift occurs at the HF level (from 0.58 to 0.26 for PmO), which, on the other hand, means that, in this way, mean values (and thus typical errors) become larger for HF. The comparably weak performance of CD, in particular at the HF level, is due to the lack of diffuse *p*-functions.

Structure parameters were investigated for those diatomic systems that turned out as worst cases in any of the previous treatments (CeF, Gd<sub>2</sub>, Lu<sub>2</sub>, NdO, PmF, PmO, SmF, TmCl), see Figure 3. Reference results were calculated by numerical gradients, as derivatives of *i*-functions are not yet available to us. The same is true for all derivatives at the MP2 level within the frozen-core approximation. Worst cases here usually are the dimers, with errors of −11 pm to +7 pm for double- $\zeta$  valence bases, −2 pm to +2 pm for triple- $\zeta$  valence bases, and <1 pm for quadruple- $\zeta$  valence bases. Compared to “usual” compounds, for metal dimers, the minima of the potential curves are rather shallow, so small inaccuracies in the description of the potential curve cause large changes in distances. Therefore, omitting the two dimers from the statistics may lead to a characterization of higher practical relevance, showing errors



**Figure 3.** Deviations (SV(P) to QZVPP bases and bases from ref 15, “CD”) to the reference for equilibrium distances of all diatomic compounds that occur as ‘worst cases’ in Figures 1 and 2. Each compound is represented by a square (DFT(BP86)), an open circle (HF), or a bar (MP2). Systems with largest deviations in distances are listed explicitly.

typically amounting to 2–3 pm for double- $\zeta$  valence bases, 0.5–0.8 pm for triple- $\zeta$  valence bases, and ca. 0.2–0.3 pm for quadruple- $\zeta$  valence bases. This probably still is a conservative estimation, since, only the worst cases were considered here. Note that, in contrast to bond energies, MP2 is not significantly more critical than HF or DFT in case of bond lengths.

#### IV. APPROXIMATIONS: RI-J AND RI-JK AUXILIARY BASES, SEMINUMERIC EXCHANGE CALCULATION

Density fitting (or resolution of the identity) methods involving preoptimized auxiliary basis sets are a proven tool to speedup calculations, in particular for the Coulomb part in DFT treatments (RI-J), but also for the exchange part in HF treatments (RI-K). This is also true for post-HF treatments (RI-MP2, RI-CC2), but, whereas for RI-K and RI-J, auxiliary basis set requirements do not significantly depend on the (orbital) basis set, as only densities resulting from occupied orbitals are to be fitted; for post-HF treatments, matters are more complicated. Here, we restrict our efforts to RI-J and RI-

K. These auxiliary basis sets were fitted at the monohydrides, exactly in the same way as done previously for the s/p/d elements. The resulting contraction patterns are listed in Table 1, and all auxiliary basis set parameters are listed in Tables S13 and S14 in the Supporting Information. RI-J auxiliary bases are about twice as large as triple- $\zeta$  valence orbital bases; RI-JK auxiliary bases are significantly larger, more than four times as large as triple- $\zeta$  valence orbital bases. RI-J errors, i.e., differences in total energies between converged RI-J-approximated and converged nonapproximated DFT(BP86) calculations (see the upper part of Table 3), amount to  $\mu \pm \sigma = 30 \pm 20 \mu E_H$ , which is somewhat larger than that observed for the lighter elements, but nearly two (three) orders of magnitude below the QZVPP orbital basis set error for HF(DFT) total energies. Note that the RI-J errors are almost the same for all orbital bases. RI-JK bases show errors of  $70 \pm 30 \mu E_H$ , which is still more than an order of magnitude below QZVPP orbital basis set error for HF total energies. For bond energies (see the lower part of Table 3), maximum errors amount to ca. 0.1 kJ/mol (RI-JK) or even less (RI-J), which is again more than an order of magnitude below the QZVPP orbital basis set errors. Speedups for RI were widely discussed previously. We restrict our considerations to  $Lu_2N$ , which is the largest molecule in our test set. Without RI, a single iteration takes ca. 8 s with TZVP bases (approximately the same for HF and DFT) and ca. 30 s with QZVPP bases. Applying RI-J in the Coulomb part reduces these times, in the case of DFT, to ca. 1 s for TZVP and 2.5 s for QZVPP. In the case of HF, matters are less favorable. For QZVPP bases, a single iteration takes 18 s; for TZVP bases with 9 s, no speedup is achieved. This is expected to be better for compounds not dominated by lanthanide atoms, because, for lighter atoms, the ratio of RI-JK auxiliary and orbital bases is smaller. For HF calculations of lanthanide-dominated compounds, the seminumeric integration scheme<sup>21</sup> for the exchange part, in combination with RI-J, turns out to work efficiently and thus might be a reasonable alternative to exact evaluation: with the smallest of our grids (m1), it takes ca. 3 s per iteration, with the “save” midsize-grid (m3) taking 6 s. The respective errors in bond energies (TZVP) with respect to the nonapproximated treatment amount to 0.6 (0.2) kJ/(mol atom) for grid m1 (m3). These errors become larger for partially filled f shells, e.g., for  $DyF_2$  ( $f^9$ ), one gets 1.6(0.8) kJ/(mol atom), which is not much smaller than the basis set error of QZVP/PP.

**Table 3.** Statistical Data of Errors Introduced by RI-J and RI-JK Approximations<sup>a</sup>

	RI-J				RI-JK			
	avg value, $\mu$	std dev, $\sigma$	worst <sup>b</sup>	at <sup>c</sup>	avg value, $\mu$	std dev, $\sigma$	worst <sup>b</sup>	at <sup>c</sup>
<b>RI Errors in Total Atomic Energies (<math>\mu E_H</math>)</b>								
QZVPP	−29	23	−81	Tb	69	32	117	Tb
QZVP	−29	23	−85	Tb	67	30	113	Tb
TZVPP	−28	18	−65	Ce	67	29	110	Tb
TZVP	−28	17	−65	Ce	60	29	106	Tb
<b>RI Errors in Bond Energies Per Atom (kJ/mol)</b>								
QZVPP	−0.005	0.024	0.046	$EuF_2$	−0.036	0.031	−0.111	$LuI_3$
QZVP	−0.005	0.024	0.046	$EuF_2$	−0.036	0.032	−0.110	$LuI_3$
TZVPP	−0.002	0.024	0.062	$EuF_2$	−0.030	0.031	−0.098	$LuI_3$
TZVP	−0.002	0.024	0.062	$EuF_2$	−0.029	0.034	0.118	PmF

<sup>a</sup>The upper part of the table shows the errors (i.e., differences of converged approximated to converged non-approximated treatments) in total atomic energies (in  $\mu E_H$ ). The lower part of the table gives errors in bond energies per atom (in kJ/mol). <sup>b</sup>The term “worst” denotes the largest error. <sup>c</sup>The term “at” denotes the atom/molecule for which the largest error occurs.

## V. SUMMARY

We presented basis sets with balanced errors of double- $\zeta$  to quadruple- $\zeta$  valence quality with two different sets of polarization functions for each level of quality, def2-SV(P) to def2-QZVPP. Bases provided by Cao and Dolg (CD) were shown to have similar errors as the triple- $\zeta$  valence bases presented here (except for the calculation of dipole moments) and served as an excellent starting point. Resulting errors for lanthanide compounds obtained with the def2 sets presented here are very similar to that of s/p/d-compounds calculated with the respective def2 bases presented previously. Thus, also recommendations for their use are the same as given previously: For energy calculations at the DFT level, SV(P) basis sets usually give more than qualitatively correct results, TZVP bases yield energies that are typically close to the basis set limit, which is practically reached with QZVP bases. To achieve similar quality at the HF level, the corresponding higher polarized sets (SVP/TZVPP/QZVPP) are needed. At the MP2 level, TZVPP (QZVPP) bases yield results of similar quality as SVP (TZVPP) bases at the HF level. Errors for other post-HF treatments are expected to be similar. For dipole moments, trends are similar, but DFT with double- $\zeta$  valence bases shows a systematic underestimation (by ca. 0.4 D). Structure parameters are less sensitive based on basis set size. Mainly independent of the level of calculation, double- $\zeta$  valence bases yield errors of 2–3 pm, which usually is sufficient for comparison with experimental data and, by far, is smaller than the usual differences between different levels of calculation. For triple- $\zeta$  valence bases, errors are <1 pm. Bonds between lanthanides show larger errors; here, instead of SV(P)/P and TZVP/PP bases, TZVP/PP and QZVP/PP bases must be used for the same errors.

All bases and auxiliary bases are available as Supporting Information, as well as via the Internet at <http://www.cosmologic.de/basis-sets/basissets.php>.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information contains the atomic energies of the bases presented here (Table S1) coordinates of the molecules of the test set (Table S2), individual bond energies (Tables S3–S5) and their statistics (Table S6), dipole moments (Tables S7 and S8) plus statistics (Table S9), individual results for distances (Table S10) and for RI errors (Table S11), further all basis set and auxiliary basis set parameters (Tables S12–S14), and the reference bases (Table S15). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [florian.weigend@kit.edu](mailto:florian.weigend@kit.edu).

### Notes

The authors declare no competing financial interest.

## ■ REFERENCES

- (1) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (2) (a) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577. (b) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835. (c) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119–124. (d) Weigend, F.; Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *119*, 12753.

- (3) Andrae, D.; Häußermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123–141.
- (4) (a) Metz, B.; Stoll, H.; Dolg, M. *J. Chem. Phys.* **2000**, *113*, 2563–2569. (b) Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. *J. Chem. Phys.* **2003**, *119*, 11113.
- (5) (a) Peterson, K. A.; Figgen, D.; Dolg, M.; Stoll, H. *J. Chem. Phys.* **2007**, *126*, 124101. (b) Figgen, D.; Peterson, K. A.; Dolg, M.; Stoll, H. *J. Chem. Phys.* **2009**, *130*, 164108.
- (6) Weigend, F.; Baltes, A. *J. Chem. Phys.* **2010**, *133*, 174102.
- (7) Roos, B. O.; Lindh, R.; Malmqvist, P.-Å.; Veryazov, V.; Widmark, P.-O. *Phys. Chem. A* **2008**, *112*, 11431–11435.
- (8) Sekiya, M.; Noro, T.; Koga, T. *Theor. Chem. Acc.* **2012**, *131*, 1247.
- (9) Pantatzis, A. D.; Neese, F. *J. Chem. Theory Comput.* **2009**, *9*, 2229–2238.
- (10) Dolg, M. *J. Chem. Theory Comput.* **2011**, *7*, 3131–3142.
- (11) Ross, R. B.; Gayen, S.; Ermler, W. C. *J. Chem. Phys.* **1994**, *100*, 8145–8155.
- (12) (a) Dolg, M.; Stoll, H.; Savin, A.; Preuss, H. *Theor. Chim. Acta* **1989**, *75*, 173–194. (b) Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1993**, *85*, 441–458.
- (13) Dolg, M.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1989**, *90*, 1730–1734.
- (14) Cao, X.; Dolg, M. *J. Chem. Phys.* **2001**, *115*, 7348–7355.
- (15) Cao, X.; Dolg, M. *J. Mol. Struct. THEOCHEM* **2002**, *581*, 139–147.
- (16) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (17) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652. (b) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822–8824.
- (18) Dunning, T. H., Jr.; Peterson, K. A.; Wilson, A. K. *J. Chem. Phys.* **2001**, *114*, 9244–9253.
- (19) Nava, P.; Sierka, M.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3372–3381.
- (20) Wilson, A. K.; Woon, D. E.; Peterson, K. A.; Dunning, T. H., Jr. *J. Chem. Phys.* **1999**, *110*, 7667–7676.
- (21) (a) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. *Chem. Phys.* **2009**, *356*, 98–109. (b) Plessow, P.; Weigend, F. *J. Comput. Chem.* **2012**, *33*, 810–816.