



# Atomic orbital basis sets

Frank Jensen\*

Electronic structure methods for molecular systems rely heavily on using basis sets composed of Gaussian functions for representing the molecular orbitals. A number of hierarchical basis sets have been proposed over the last two decades, and they have enabled systematic approaches to assessing and controlling the errors due to incomplete basis sets. We outline some of the principles for constructing basis sets, and compare the compositions of eight families of basis sets that are available in several different qualities and for a reasonable number of elements in the periodic table. © 2012 John Wiley & Sons, Ltd.

How to cite this article:

*WIREs Comput Mol Sci* 2013, 3: 273–295 doi: 10.1002/wcms.1123

## INTRODUCTION

The use of basis sets for expanding the molecular orbitals in wave function or Kohn–Sham density functional methods is an essential component of contemporary methods for describing the electronic structure of molecular and extended systems. Nuclear-centered Slater-<sup>1</sup> or Gaussian-type<sup>2</sup> functions have dominated for molecular systems, with Gaussian functions being preferred due to the better computational efficiency. Plane-wave basis functions are often used for extended systems, as they are naturally suited to periodic boundary conditions; but for any reasonable number of plane waves, this necessitates the use of a pseudo-potential for representing the atomic core electrons/potential. Nuclear-centered basis functions can also be used for periodic systems, and this treats all electrons on an equal footing. Recent developments have investigated the use of finite-element methods where piecewise polynomials are used for representing the orbitals.<sup>3–5</sup>

The goal of a basis set is to provide the best representation of the unknown molecular orbitals (or electron density), with as small a computational cost as possible. Because different theoretical methods and molecular properties have different basis set demands, different computer architectures and algorithms have different efficiency requirements, and the desired accuracy varies with the application, it is not possible to design one ‘optimum’ basis set. Indeed, the large

number of different basis sets proposed over the years is a testament to these conflicting demands.

In this review, we will summarize some of the principles for constructing and classifying basis sets, with focus on modern hierarchical basis sets. Only basis sets composed of nuclear-centered Gaussian-type functions will be discussed, but many of the principles and conclusions hold for Slater-type functions as well. We will concentrate on basis sets for the first 36 atoms in the periodic table [hydrogen (H) to krypton (Kr)] as many of the popular basis sets are only available for these elements, but the trends and principles carry over to elements in the remaining part of the periodic table as well.

All-electron calculations for systems containing atoms from the lower part of the periodic table must include relativistic effects for accurate results, and this leads to some differences in the basis set requirements compared to nonrelativistic methods. We will not discuss basis sets for relativistic methods in this review. A significant computational saving can be obtained for systems with many-electron atoms by replacing the core electrons by a pseudo-potential or model potential, and this can also to some extent account for relativistic effects. The use of pseudo-potentials has been the subject of recent reviews by Dolg and Cao, and these can be consulted for further information.<sup>6,7</sup> A discussion of plane wave or finite-element methods, and the topic of auxiliary basis sets for density fitting, is beyond the scope of the present review.

We will focus on eight families of basis sets that are available in several quality levels and defined for a reasonable number of elements in the periodic table:

\*Correspondence to: frj@chem.au.dk

Department of Chemistry, Aarhus University, Aarhus, Denmark  
DOI: 10.1002/wcms.1123

1. The Pople-style *k-lmn*G basis sets.<sup>8–20</sup>
2. The Ahlrichs SVP, TZP, QZP basis sets in their Def2 version.<sup>21</sup>
3. The XZP basis sets developed by Jorge and coworkers.<sup>22–26</sup>
4. The Sapporo basis sets developed by Koga and coworkers.<sup>27–30</sup>
5. The ANO basis sets developed by Roos and coworkers.<sup>31–35</sup>
6. The cc-pVXZ basis sets developed by Dunning<sup>36</sup> and Peterson and coworkers.<sup>37–39</sup>
7. The *n*ZaP basis sets developed by Petersson and coworkers.<sup>40–42</sup>
8. The pc-*n* basis sets developed by Jensen and coworkers.<sup>43–48</sup>

## PRIMITIVE AND CONTRACTED BASIS FUNCTIONS

An unknown one-electron function, such as a molecular orbital  $\phi$ , can be expanded in a set of known functions  $\chi$ , the basis set<sup>49,50</sup>:

$$\phi = \sum_{i=1}^M c_i \chi_i. \quad (1)$$

The coefficients  $c_i$  are in Hartree–Fock (HF) and Kohn–Sham density function theory (DFT) determined by minimizing the total energy, which by traditional methods lead to a matrix eigenvalue problem that is solved iteratively to provide a self-consistent field (SCF) solution. The matrix elements contain multidimensional integrals over basis functions, of which those involving the electron–electron interaction (two-electron integrals) completely dominate the computational effort. We will, in the present context, not be concerned with determination of the molecular expansion coefficients, but focus on the basis functions, which will be taken as Cartesian Gaussian-type functions:

$$\chi_i = N(x - X)^k (y - Y)^l (z - Z)^m e^{-\zeta_i(r-R)^2}. \quad (2)$$

The center of such a *primitive* function is  $R(X, Y, Z)$ , typically a nuclear position, the sum of  $k$ ,  $l$ , and  $m$  defines the angular momentum (e.g.,  $k + l + m = 1$  is a p-function),  $\zeta_i$  is the exponent providing the radial extent of the function, and  $N$  is a normalization constant. The accuracy of the expansion in Eq. (1) is determined by the number of functions  $M$ , their distribution in terms of angular momentum, and the values of the exponents  $\zeta_i$  [Eq. (2)].

The primitive basis set is for computational reasons usually *contracted*, by forming  $K$  fixed linear combinations  $\kappa_j$  from  $M$  primitive functions  $\chi_i$  ( $K < M$ ):

$$\begin{aligned} \kappa_j &= \sum_{i=1}^M d_{ij} \chi_i, \\ \phi &= \sum_{j=1}^K c_j \kappa_j. \end{aligned} \quad (3)$$

The physical reason for contraction is that the core orbitals are insensitive to the molecular environment, and can therefore be described by a fixed linear combination with little loss of accuracy. Contraction reduces the number of coefficients  $c_j$  from  $M$  to  $K$  for each molecular orbital, and thus simplifies the variational problem. The number of two-electron integrals that need to be calculated is unaffected by contraction, although the efficiency of the computational implementation may depend on the exact nature of the contraction.

The typical notation used to indicate the reduction in the number of variational parameters is with parenthesis for the primitive set and square brackets for the contracted set, e.g., (9s4p)  $\rightarrow$  [3s2p] denotes nine primitive s-functions contracted to three s-functions, and four primitive p-functions contracted to two p-functions. This does not specify *how* the contraction is done, only the *number* of resulting functions. The explicit contraction scheme can be given in curly brackets as {621/31} which indicates that the first, second, and third contracted s-functions are constructed of six, two, and one primitive functions, respectively, whereas the two contracted p-functions contain three and one primitive functions.

Two types of contraction can be used, known as *segmented*<sup>51</sup> and *general*<sup>52</sup> contraction. In the purest versions, a segmented contraction forms a disjoint set of the primitive functions (i.e., a given primitive function  $\chi_i$  has only a nonzero contraction coefficient  $d_{ij}$  in one of the contracted functions  $\kappa_j$ ), whereas a general contraction uses *all* the primitive functions in *all* the contracted functions (i.e., all  $d_{ij} \neq 0$ ).

Contraction is in practice rarely completely segmented or general. A general contraction can be made partly segmented by a simple mathematical transformation, which is equivalent to the variable elimination used in solving linear equations.<sup>53</sup> This transformation results in each of the  $K$  contracted functions having only nonzero contributions from  $M - K + 1$  primitive functions. This transformation is exact, i.e., the same results are obtained with the untransformed and transformed basis sets. The

transformed basis set can be further segmented by neglecting small coefficients (e.g., less than  $10^{-5}$ ), but this represents a real, albeit small, reduction of the accuracy by which the molecular orbitals can be represented.

A set of general contracted functions can for atomic systems at the HF or DFT level generate the same result as the full set of primitive functions, provided that the contraction coefficients are taken as the SCF (HF or DFT) coefficients. For molecular systems and for correlated methods, a general contraction will always involve a penalty, with the magnitude depending on the level of contraction, i.e., the number of contracted functions  $K$ .

A segmented contraction always reduces the accuracy by which the orbitals can be represented, even for atoms. The inner primitive functions (large  $\zeta$ ) make only a large contribution to the 1s-orbital, and the outer primitive functions (small  $\zeta$ ) are usually left uncontracted (i.e., a contracted function involving only one primitive function), because these describe the environmentally sensitive valence orbitals. The primitive functions with intermediate exponents, however, are important for describing both core and valence orbitals (or orbitals from different shells in the general case). Restricting them to participate in only one contracted function can therefore give a significant reduction in accuracy of the final set of functions. It is therefore a common practice to *duplicate* one or more functions in the intermediate exponent range, which can be viewed as a step toward a general contraction. In a general contraction, on the other hand, the basis functions describing the outer valence part of the orbitals are normally left uncontracted, i.e., a partly segmented contraction. The terms segmented and general contraction should thus be taken as limiting cases, with actual basis sets containing elements of both contraction forms.

Segmented contraction has traditionally been the preferred method, and most integral codes in electronic structure programs have been written with such basis sets in mind. Although these programs can also treat general contracted basis sets, they do this by replicating the primitive functions as many times as they enter the contracted functions, thereby significantly increasing the number of primitive basis functions (potentially from  $M$  to  $M \times K$ ) and leading to computational inefficiency. Some programs have been written with general contracted basis sets in mind, and they can handle segmented basis sets with no loss in efficiency, because segmentation is just a special case of general contraction.

An additional factor influencing the computational efficiency is the use of integral screen-

ing to avoid calculating some of the two-electron integrals.<sup>54–56</sup> This screening becomes efficient for large systems and works most favorably with small basis sets, and it is the premise for reducing the HF/DFT scaling of the computational time with system size from formally  $N^4$  to  $\sim N^2$  without employing special techniques. The screening relies on calculating an upper limit for each batch of two-electron integrals (a batch is a group of functions sharing the same exponent), and if this upper limit is below a suitable threshold, the entire batch of integrals can be neglected. The screening is performed at the contracted basis set level, and therefore works most efficiently for segmented basis sets. If an integral batch involving a specific set of contracted functions can be neglected, all of the integrals involving the underlying primitive functions can be neglected. In a general contraction, on the other hand, all primitives may contribute to all contracted functions, and screening is therefore less efficient.

## BASIS SET CLASSIFICATION

Basis sets can be described by the number of primitive functions and the contraction of these functions. For classification purposes, however, it is useful to focus on the number of contracted functions, since this indicates the flexibility of the basis set to describe different molecular environments. Except for very large basis sets, the errors in the absolute energy due to basis set deficiencies are large, and obtaining accurate relative energies therefore relies on substantial error cancellation. This means that basis set balance is of utmost importance, both internally in terms of number of functions with different angular momentum, and also externally between different elements.<sup>57,58</sup>

A *minimum* or *single zeta* (SZ) basis set has the number of contracted functions equal to the number of occupied orbitals for the isolated atom, with the addendum that all components of a given function type are included. For hydrogen, carbon, and sulfur, this means basis sets of 1s, 2s1p, and 3s2p, respectively. This can provide the basis set limiting energy for isolated atoms at HF and DFT levels, if a large number of primitive functions are used in the contraction. SZ basis sets, however, provide poor results for molecules, since they are inadequate at describing the deviation from the spherical symmetry of the atoms and the process of bond formation. For methods including electron correlation, where the virtual (unoccupied) orbitals are used for expanding the many-electron wave function, SZ basis sets are likewise unsuitable.

The next step up in basis set quality is to use two independent functions for each atomic orbital. The function(s) describing the core orbitals are in practice rarely duplicated, thus such *double zeta* (DZ) basis sets are normally only double zeta in the valence region, i.e., 2s, 3s2p, and 4s3p for hydrogen, carbon, and sulfur, respectively. DZ basis sets are much better at describing changes due to molecular bond formation, as functions with different exponents can be used in different ratios to describe that bonding is different in different directions. At this point, a higher angular momentum function should also be added to describe the charge polarization for HF and DFT methods, and the electron–electron cusp at correlated levels. Such basis sets are usually called *double zeta plus polarization* (DZP).

The idea of increasing the number of s- and p-functions, and at the same time adding higher angular momentum functions, can be continued to produce *triple zeta plus polarization* (TZP), *quadruple zeta plus polarization* (QZP), etc. types of basis sets. In early work, the TZ and QZ notations directly reflected the number of s- and p-functions, whereas higher-order polarization functions were added in an *ad hoc* manner. The seminal work of Dunning,<sup>36</sup> based on the concept of atomic natural orbitals,<sup>59</sup> strongly suggested that polarization functions should be added in well-defined sequences, such that the total basis set error is approximately equally distributed between all types of basis functions. This introduced the important concept that the basis set composition in terms of number of functions of different angular momentum can be determined by analyzing energy contributions, rather than relying on subjective criteria, such as performance for a selection of molecular systems and properties. The XZP nomenclature in modern classifications therefore normally reflects the highest angular momentum function included in the basis set, and this will also be the usage in the present review. A DZP basis set thus indicates that it includes a function with one higher angular momentum than required by the occupied (valence) atomic orbitals, a TZP basis set includes a function with two higher angular momenta than required by the occupied atomic orbitals, and so on.<sup>60</sup>

The XZP classification scheme serves as a guideline for the quality of a basis set, and in favorable cases provides a systematic convergence toward the basis set limit. Tables 1–4 show the composition for first-, second-, and third-row s-, p-, and d-block elements in terms of primitive and contracted functions for the eight families of basis sets mentioned in the introduction.

The XZP classification is unambiguous for p-block elements from the first two rows in the periodic table, and the contracted basis set compositions are very similar for the eight families (Tables 1 and 2). A choice regarding polarization functions must be made for s-block elements. As they only have occupied (valence) s-orbitals, p-functions are formally the first type of polarization. The low-lying unoccupied p-orbitals, however, often play an important role in the bonding of s-block elements, and d-functions may thus also be considered as the first type of polarization function. Different families of basis sets make different choices, and in some cases also different choices between first-, second-, and third-row s-block elements (Tables 1–3). The situation is similarly ambiguous for the third-row p-elements, where the atoms have occupied 3d-orbitals. If the 3d-orbitals are considered part of the atomic core and not taking part in the chemical bonding, then a d-function is the first polarization function. If, on the other hand, the 3d-electrons are considered part of the valence electrons, then an f-function is the first polarization function. Different families of basis sets again make different choices (Table 3). The molecular bonding in third-row transition metal systems is dominated by the 3d-orbitals, and almost all basis set families consider f-functions to be the first polarization function (Table 4).

Although the XZP classification is useful for judging the quality of the results, it should be noted that details of how, e.g., a TZP basis set is constructed will influence the accuracy obtained, and this is the topic of the next sections.

## PRIMITIVE BASIS SETS

Isolated atoms with  $Z \leq 20$  (Ca) can at HF and DFT levels be described completely by s- and p-functions only, atoms up to  $Z \leq 56$  (Ba) by s-, p-, and d-functions only, whereas the remaining atoms require also f-functions. For molecular systems, and for methods involving electron correlation, angular momentum functions higher than sp, spd, and spdf, respectively, are required to describe polarization and correlation effects. By far the most common procedure is to determine the basis functions describing the occupied atomic orbitals by HF/DFT optimization at the atomic level, and treating the question of polarization functions subsequently.

## Optimized Primitive Basis Sets

The best possible uncontracted basis set in an energetic sense can be constructed by minimizing the

**TABLE 1** | Basis Set Compositions for First-Row s- and p-Block Atoms

Quality			s-Block Elements										p-Block Elements													
			Primitives					Contracted					Primitives					Contracted								
			s	p	d	f	g	h	s	p	d	f	g	h	s	p	d	f	g	h						
SZ	Pople	STO-3G	6	3				2	1				6	3				2	1							
	Pople	3-21G	6	3				3	2				6	3				3	2							
DZ	Jensen	pc-0	5	1				3	1				5	3				3	2							
	Pople	6-31G*	10	4	1			3	2	1			10	4	1			3	2	1						
DZP	Ahlrichs	Def2-SVP	7	4				3	2				7	4	1			3	2	1						
	Jorge	DZP	9	5	1			4	2	1			9	5	1			4	2	1						
	Sapporo	DZP	8	4				4	2				9	4	2			3	2	1						
	Roos	ANO-DZP	14	9	4			4	3	1			14	9	4			3	2	1						
	Dunning	cc-pVDZ	9	4	1			3	2	1			9	4	1			3	2	1						
	Petersson	2ZaP	10	5	1			4	3	1			11	8	1			4	3	1						
	Jensen	pc-1	7	3				3	2				7	4	1			3	2	1						
	Pople	6-311G(2df)	11	5	2	1		4	3	2	1		11	5	2	1		4	3	2	1					
	Ahlrichs	Def2-TZVP	11	4	1			5	3	1			11	6	2	1		5	3	2	1					
	Jorge	TZP	10	6	2	1		5	3	2	1		10	6	2	1		5	3	2	1					
TZP	Sapporo	TZP	9	4	3			6	4	2			10	5	3	2		4	3	2	1					
	Roos	ANO-TZP	14	9	4	3		5	4	2	1		14	9	4	3		4	3	2	1					
	Dunning	cc-pVTZ	11	5	2	1		4	3	2	1		10	5	2	1		4	3	2	1					
	Petersson	3ZaP	13	8	2	1		5	4	2	1		14	11	2	1		5	4	2	1					
	Jensen	pc-2	10	4	1			4	2	1			10	6	2	1		4	3	2	1					
	Ahlrichs	Def2-QZVP	15	7	2	1		7	4	2	1		15	8	3	2	1	7	4	3	2	1				
	Jorge	QZP	11	7	3	2	1	6	4	3	2	1	11	7	3	2	1	6	4	3	2	1				
	Sapporo	QZP	9	6	4	1		7	5	3	1		12	7	4	3	2	5	4	3	2	1				
	Roos	ANO-QZP	14	9	4	3		6	5	3	2		14	9	4	3	2	5	4	3	2	1				
	Dunning	cc-pVQZ	12	6	3	2	1	5	4	3	2	1	12	6	3	2	1	5	4	3	2	1				
QZP	Petersson	4ZaP	14	6	2	1		6	3	2	1		17	14	3	2	1	6	5	3	2	1				
	Jensen	pc-3	14	6	2	1		6	3	2	1		14	9	4	2	1	6	5	4	2	1				
	Jorge	5ZP	14	8	4	3	2	1	7	5	4	3	2	1	14	8	4	3	2	1	7	5	4	3	2	1
	Dunning	cc-pV5Z	14	8	4	3	2	1	6	5	4	3	2	1	14	8	4	3	2	1	6	5	4	3	2	1
	Petersson	5ZaP	19	8	3	2	1		8	4	3	2	1	20	16	4	3	2	1	7	6	4	3	2	1	
	Jensen	pc-4	19	8	3	2	1		8	4	3	2	1	18	11	6	3	2	1	8	7	6	3	2	1	
5ZP	Jensen	pc-3	14	6	2	1		6	3	2	1		14	9	4	2	1	6	5	4	2	1				
	Jorge	5ZP	14	8	4	3	2	1	7	5	4	3	2	1	14	8	4	3	2	1	7	5	4	3	2	1
	Dunning	cc-pV5Z	14	8	4	3	2	1	6	5	4	3	2	1	14	8	4	3	2	1	6	5	4	3	2	1
	Petersson	5ZaP	19	8	3	2	1		8	4	3	2	1	20	16	4	3	2	1	7	6	4	3	2	1	

TABLE 2 | Basis Set Compositions for Second-Row s- and p-Block Atoms

Quality			s-Block Elements										p-Block Elements									
			Primitives					Contracted					Primitives					Contracted				
			s	p	d	f	g	h	s	p	d	f	g	h	s	p	d	f	g	h		
SZ	Pople	STO-3G	9	6				3	2				9	6							3	2
	Pople	3-21G	9	6				4	3				9	6							4	3
DZ	Jensen	pc-0	8	5				4	2				8	6							4	3
	Pople	6-31G*	16	10	1			4	3	1			16	10	1						4	3
DZP	Ahlrichs	Def2-SVP	10	7	1			4	3	1			10	7	1						4	3
	Jorge	DZP	12	9	1			7	4	1			12	9	1						7	4
	Sapporo	DZP	11	7	2			5	4	1			12	9	3						4	3
	Roos	ANO-DZP	17	12	6			5	2	2			17	12	5						4	3
	Dunning	cc-pVDZ	12	8	1			4	3	1			12	8	2						4	3
	Petersson	2Zap	15	9	1			6	4	1			14	10	2						5	4
	Jensen	pc-1	11	7				4	2				11	8	1						4	3
	Pople	6-311G(2df)	13	9	2	1		6	5	2	1		12	9	2	1					6	5
TZP	Ahlrichs	Def2-TZVP	14	8	3			5	4	3			14	9	3	1					5	5
	Jorge	TZP	13	10	2	1		8	5	2	1		13	10	2	1					8	5
	Sapporo	TZP	15	9	4	1		7	6	3	1		13	9	4	2					5	4
	Roos	ANO-TZP	17	12	6			6	3	2			17	12	5	4					5	4
	Dunning	cc-pVTZ	15	10	2	1		5	4	2	1		15	9	3	1					5	4
	Petersson	3Zap	17	12	2	1		6	5	2	1		17	13	3	1					6	5
	Jensen	pc-2	13	9	1			5	3	1			13	10	2	1					5	4
	Ahlrichs	Def2-QZVP	20	12	4	1		9	5	4	1		20	14	4	2	1				9	6
QZP	Jorge	QZP	14	11	3	2	1	9	6	3	2	1	14	11	3	2	1				9	6
	Sapporo	QZP	16	10	5	3	1	8	7	5	3	1	16	11	5	2	2				6	5
	Roos	ANO-QZP	17	12	6	2		7	4	4	2		17	12	5	4	2	2			6	5
	Dunning	cc-pVQZ	16	12	3	2	1	6	5	3	2	1	16	11	4	2	1				6	5
	Petersson	4Zap											20	16	4	2	1				7	6
	Jensen	pc-3	17	12	2	1		6	4	2	1		17	13	4	2	1				6	5
5ZP	Jorge	5ZP	18	12	4	3	2	1	10	7	4	3	2	1	18	12	4	3	2	1	10	7
	Dunning	cc-pV5Z	20	14	4	3	2	1	7	6	4	3	2	1	20	12	5	3	2	1	7	6
	Petersson	5Zap											23	19	5	3	2	1			8	7
	Jensen	pc-4	21	15	3	2	1	7	5	3	2	1	21	16	6	3	2	1			7	6

**TABLE 3** | Basis Set Compositions for Third-Row s- and p-Block Atoms

Quality	Label	Name	s-Block Elements										p-Block Elements													
			Primitives					Contracted					Primitives					Contracted								
			s	p	d	f	g	h	s	p	d	f	g	h	s	p	d	f	g	h	s	p	d	f	g	h
SZ	Pople	STO-3G	12	9				4	3					12	9	3				4	3	1				
	Pople	3-21G	12	9				5	4					12	9	3				5	4	1				
DZ	Jensen	pc-0	8	7				5	3					8	7	4				5	4	1				
	Pople <sup>1</sup>	6-31G*	22	16	4			5	4	2				22	16	5				5	4	3				
DZP	Pople <sup>1</sup>	6-31G*	22	16	1			5	4	1				14	11	6				6	4	2				
	Ahlrichs	Def2-SVP	14	9	4			5	3	2				14	10	6				5	4	3				
	Jorge	DZP	15	12	1			8	6	1				15	12	5	1			8	6	2	1			
	Sapporo	DZP	16	12	2			6	5	1				17	14	9				5	4	2				
	Roos	ANO-DZP	20	16	6			6	3	2				20	17	11	4			5	4	2	1			
	Dunning	cc-pVDZ	14	11	5			5	4	2				14	11	6				5	4	2				
	Petersson	2ZaP	17	12	1			6	5	1				18	13	7				6	5	2				
	Jensen	pc-1	12	9	4			5	3	2				12	10	7				5	4	2				
TZP	Pople	6-311G(2df)	14	11	5	1		8	7	3	1			15	13	7	1			8	7	4	1			
	Ahlrichs	Def2-TZVP	17	12	4			6	5	3				17	13	8	1			6	5	4	1			
	Jorge	TZP	16	13	4	1		9	6	4	1			16	13	6	2	1		9	6	3	2	1		
	Sapporo	TZP	18	13	4	1		8	6	3	1			18	14	10	2			6	5	3	1			
	Roos	ANO-TZP	20	16	6			7	4	2				20	17	11	4	2		6	5	3	2	1		
	Dunning	cc-pVTZ	20	14	6	1		6	5	3	1			20	13	9	1			6	5	3	1			
	Petersson	3ZaP	20	14	2	1		7	6	2	1			21	16	10	1			7	6	3	1			
	Jensen	pc-2	16	11	5			6	4	3				16	13	9	1			6	5	3	1			
QZP	Ahlrichs	Def2-QZVP	24	18	6	1		11	6	4	1			24	20	10	2	1		11	7	4	2	1		
	Ahlrichs	Def2-QZVPP	24	18	6	3		11	6	4	3			24	20	10	4	1		11	7	4	4	1		
	Sapporo	QZP	19	14	5	3	1	9	7	4	2	1		19	14	10	2	2		7	6	4	2	1		
	Roos	ANO-QZP	20	16	6	4		8	5	4	2			20	17	11	4	2		7	6	4	3	2		
	Dunning	cc-pVQZ	22	16	7	2	1	7	6	4	2	1		21	16	12	2	1		7	6	4	2	1		
	Jensen	pc-3	20	14	6	1		7	5	4	1			20	16	11	2	1		7	6	4	2	1		
5ZP	Dunning	cc-pV5Z	26	18	8	3	2	1	8	7	5	3	2	1	26	17	13	3	2	1	8	7	5	3	2	1
	Jensen	pc-4	24	17	8	2	1	8	6	5	2	1		24	19	13	4	2	1	8	7	5	4	2	1	

<sup>1</sup>Two different definitions of the 6-31G\* basis set are available.<sup>12,13</sup>



**TABLE 4** | Basis Set Compositions for Third-Row d-Block Atoms

Quality	Label	Name	Primitives							Contracted						
			s	p	d	f	g	h	i	s	p	d	f	g	h	i
SZ	Pople	STO-3G	12	9	3					4	3	1				
DZ	Pople	3-21G	12	9	3					5	4	2				
	Jensen	pc-0	8	7	4					5	3	1				
DZP	Pople	6-31G*	22	16	4	1				5	4	2	1			
	Ahlrichs	Def2-SV(P)	14	9	5					5	3	2				
	Ahlrichs	Def2-SVP	14	9	5	1				5	3	2	1			
	Jorge	DZP	15	12	5	1				8	6	2	1			
	Sapporo	DZP	16	13	9	2				6	4	3	1			
	Roos	ANO-DZP	21	15	10	6				5	4	2	1			
	Dunning	cc-pVDZ <sup>1</sup>	20	16	8	2				6	5	3	1			
	Petersson	2ZaP	17	12	7	1				6	5	2	1			
	Jensen	pc-1	12	9	6	1				5	3	2	1			
TZP	Pople	6-311G(2df)	14	9	5	2	1			9	5	3	2	1		
	Ahlrichs	Def2-TZVP	17	11	7	1				6	4	4	1			
	Ahlrichs	Def2-TZVPP	17	12	7	2	1			6	5	4	2	1		
	Jorge	TZP	16	13	6	2	1			9	6	3	2	1		
	Sapporo	TZP	18	13	10	3	2			8	6	4	2	1		
	Roos	ANO-TZP	21	15	10	6	4			6	5	3	2	1		
	Dunning	cc-pVTZ	20	16	8	2	1			7	6	4	2	1		
	Petersson	3ZaP	20	14	10	2	1			7	5	3	2	1		
	Jensen	pc-2	16	11	8	2	1			6	4	3	2	1		
QZP	Ahlrichs	Def2-QZVP	24	18	10	3	1			11	6	5	3	1		
	Ahlrichs	Def2-QZVPP	24	18	10	4	2			11	6	5	4	2		
	Sapporo	QZP	18	13	11	3	2	1		9	7	5	3	2	1	
	Roos	ANO-QZP	21	15	10	6	4			7	6	4	3	2		
	Dunning	cc-pVQZ	22	18	11	3	2	1		8	7	5	3	2	1	
	Jensen	pc-3	20	14	10	3	2	1		7	5	4	3	2	1	
5ZP	Dunning	cc-pV5Z	28	20	12	4	3	2	1	9	8	6	4	3	2	1
	Jensen	pc-4	24	17	12	5	3	2	1	8	6	5	5	3	2	1

<sup>1</sup>The cc-pVDZ is defined as a DZP contraction of the same primitive set of functions as the cc-pVTZ.

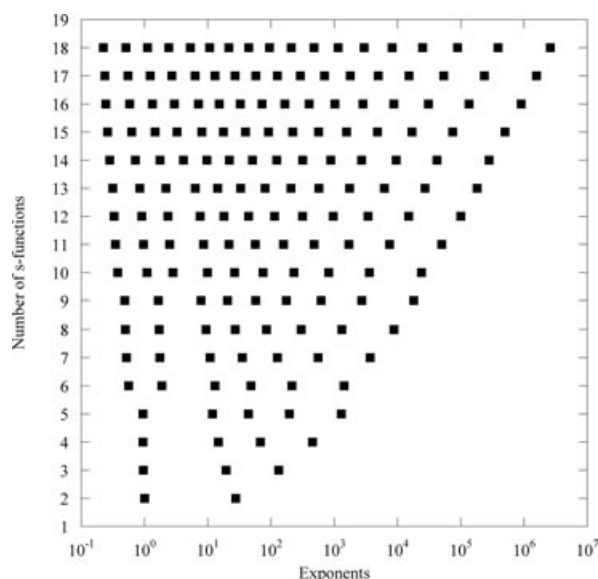
energy as a function of the exponents  $\zeta_i$  [Eq. (2)] because the expansion coefficients  $c_i$  are determined variationally by the SCF procedure [Eq. (1)]. The exponent optimization can be performed either by a pattern search involving only the energy itself,<sup>61</sup> by methods involving the analytical gradient of the energy with respect to the exponents,<sup>62,63</sup> or by methods using also the second derivative of the energy with respect to the exponents.<sup>64</sup> Most of the early work on basis set optimizations relied on nonderivative methods, whereas more recent work has used analytical exponent gradients.

Whatever the optimization method used, it will only determine a local minimum, i.e., it is possible that optimizations starting from different initial exponents may lead to different final results. A specific example has been reported by Mezey et al.,<sup>65</sup> where two different sets of s-exponents were determined

for a (6s3p) basis set for the Li–F atoms. In one of the two sets, the five inner functions primarily describe the 1s-orbital, with the single remaining function mainly describes the 2s-orbital, whereas the other set of exponents corresponds to a 4/2 split for the 1s/2s-orbitals.

The problem with multiple minima is mainly associated with small basis sets, and has a clear physical interpretation related to the orbital shell structure of the atom.<sup>46</sup> The shell effect is clearly visible in the ratio between optimized exponents in the basis set, as illustrated in Figure 1 for s-functions for the neon atom. With five s-functions, the ratio between  $\zeta_2$  and  $\zeta_1$  ( $\zeta_1$  being the smallest exponent) is significantly larger than the ratio between  $\zeta_3$  and  $\zeta_2$ . This exponent ‘gap’ reflects the nodal structure of the 2s-orbital, where a basis function with an exponent  $\sim 4$  has a maximum with a distance to the nucleus





**FIGURE 1** | Hartree-Fock optimized s-exponent values for the neon atom for basis sets of increasing size.

corresponding to the 2s-node, and is therefore energetically less important than basis functions with exponents larger or smaller than  $\sim 4$ . The shell gap is present until 14 s-functions, beyond which the exponents cover the whole range rather uniformly, and multiple minima are therefore mainly a problem for small basis sets and disappear as the number of functions increase.

### Parameterized Primitive Basis Sets

The exponents of a primitive basis set form a highly coupled nonlinear set of variables. A complete optimization, even with analytical exponent gradients, is a nontrivial task for large basis sets and often involves many hundreds of iterations. Furthermore, as the basis set approaches completeness, the energy as a function of the exponents becomes a very flat surface.

As an alternative to performing the optimization directly with the exponents as variables, the exponents can be generated by a parameterized formula, and the parameters optimized instead. Already, very early it was noted that the exponents of completely optimized basis sets had some regular features. The simplest parameterization takes the ratio between successive exponents to be constant, forming a so-called *even-tempered* sequence<sup>66</sup>:

$$\zeta_i = \alpha\beta^i \quad i = 1, 2, \dots, M. \quad (4)$$

Even-tempered basis sets have only two parameters ( $\alpha$  and  $\beta$ ) to optimize, independent of the number

of functions  $M$ . Schmidt and Ruedenberg<sup>67</sup> furthermore noted that the  $\alpha$  and  $\beta$  parameters had to obey certain rules in order for the basis set to approach completeness as  $M$  increases, and provided a suitable parameterization. The even-tempered formula allows construction of basis sets which converge toward the basis set limit for atoms, and this has been used for constructing ‘universal’ basis sets.<sup>68,69</sup> Jorge and de Castro<sup>70</sup> have used a similar concept for constructing basis sets by a generator coordinate approach.

An even-tempered basis set has the same ratio between exponents in the whole range from  $\alpha\beta$  to  $\alpha\beta^M$ , but the results from fully optimized basis sets suggest that the ratio should increase both for the largest and smallest exponents, as illustrated in Figure 1. The well-tempered basis sets achieve the former by the exponent parameterization shown in Eq. (5), where the  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  parameters are optimized for each atom<sup>71,72</sup>:

$$\zeta_i = \alpha\beta^{i-1} \left[ 1 + \gamma \left( \frac{i}{M} \right)^\delta \right] \quad i = 1, 2, \dots, M. \quad (5)$$

The even-tempered parameterization can be generalized by recognizing that the formula in Eq. (4) can be written as a linear expansion by taking the logarithm

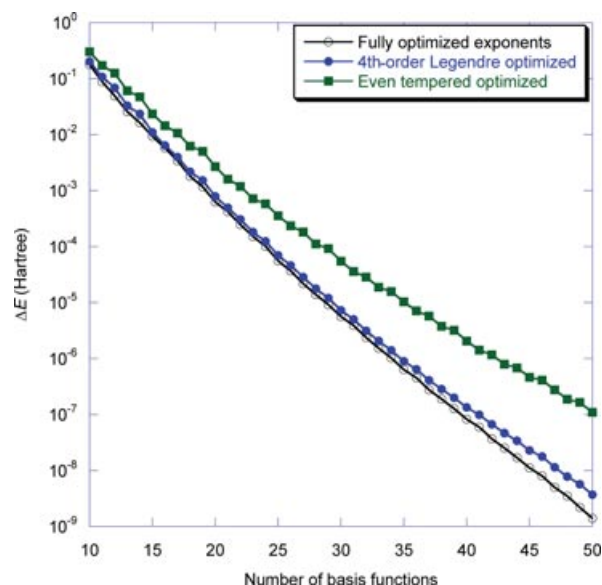
$$\ln(\zeta_i) = \ln(\alpha) + i \ln(\beta). \quad (6)$$

Petersson et al.<sup>73</sup> have suggested an exponent parameterization in terms of Legendre polynomials  $P_k$ :

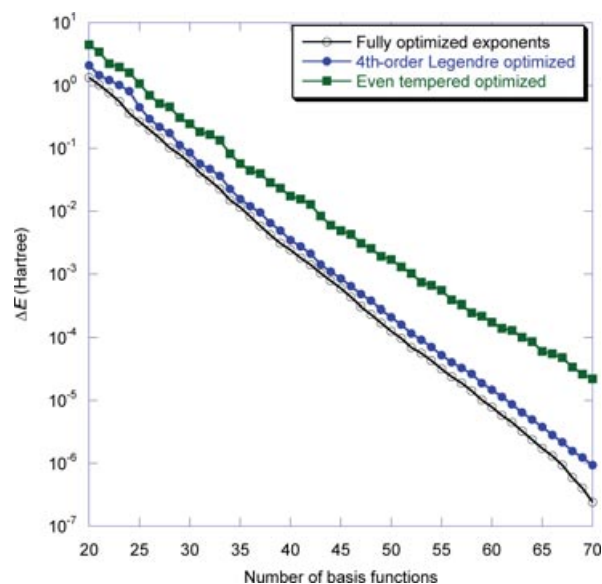
$$\ln(\zeta_i) = \sum_{k=0}^{k_{\max}} A_k P_k \left( \frac{2i-2}{M-1} - 1 \right); \quad i = 1, 2, \dots, M \quad (7)$$

A second-order expansion ( $k_{\max} = 1$ ) is equivalent to the even-tempered formula ( $A_0 = \ln(\alpha)$ ,  $A_1 = \ln(\beta)$ ), and a fourth-order expansion is able to reproduce the results from a fully optimized basis set closely.

Parameterized formulas require more functions for achieving a given accuracy than fully optimized basis sets, and for even-tempered basis sets the penalty increases with the basis set size and with the atomic number. This is illustrated in Figures 2 and 3, where the energy convergence toward the HF limit for the Ne and Kr atoms are displayed as a function of the basis set size. The additional number of functions required for an even-tempered parameterization to reach the same accuracy as a fully optimized basis set is 3 at the milli-Hartree level and 8 at the micro-Hartree level for



**FIGURE 2** | Hartree–Fock energy contribution of basis functions for the neon atom, using either a full optimization, a fourth-order Legendre [Eq. (7)] or an even-tempered [Eq. (4)] parameterization for determining the basis function exponents.



**FIGURE 3** | Hartree–Fock energy contribution of basis functions for the krypton atom, using either a full optimization, a fourth-order Legendre [Eq. (7)] or an even-tempered [Eq. (4)] parameterization for determining the basis function exponents.

Ne, whereas the corresponding values for Kr are 9 and ~20 functions. A fourth-order Legendre polynomial provides an efficient parameterization, with a penalty of typically only one or two extra functions, even for very large basis sets.

## Fitted Basis Sets

Slater-type orbitals (STO) depend exponentially on the nuclear–electron distance, rather than the Gaussian exponential square dependence:

$$\chi_{\text{STO}} = \text{Ne}^{-\xi(r-R)}. \quad (8)$$

Because STOs can represent the exact solutions for hydrogen-like atoms, they provide a better representation than Gaussian functions for many-electron systems on a function-to-function comparison. Multicenter two-electron integrals, however, are cumbersome with STOs, but are much easier with Gaussian functions. The use of density fitting methods can partly alleviate the problem of difficult multicenter integrals, but the overall computational effort appears to be similar with the two types of basis sets.<sup>74,75</sup>

The use of STOs has been an efficient method of achieving high accuracy for atoms, and such basis sets were developed very early.<sup>76</sup> For use in molecular calculations, a common practice was to represent the optimized atomic STOs by a fixed linear combination of Gaussian functions.<sup>77</sup> The optimum representation was done by fitting the exponents and coefficients to the STOs in a least squares sense, and the fitting error decreases as the number of Gaussian functions is increased. Such basis sets have been developed by Pople and coworkers<sup>78–80</sup> for a large part of the elements in the periodic table, and are known by their acronym STO-*n*G, where *n* indicates the number of Gaussian functions used in the fitting. As these are SZ-type basis sets, there is for application purposes little improvement of going beyond *n* = 3, and the STO-3G basis set is probably the most popular SZ basis set.

## POLARIZATION FUNCTIONS

The addition of higher angular momentum functions allows the molecular density to distort from the spherical symmetry of the atoms at the HF or DFT levels, and these functions for correlated calculations, in addition, serve to describe the electron correlation cusp. The accepted notation is to call these ‘polarization’ functions, despite their dual purpose. In the following, we will discuss polarization functions for first- and second-row p-block elements only, thereby avoiding the issue regarding s-block elements and whether to consider the 3d-orbitals to be core or valence orbitals for third-row p-block atoms.

Polarization functions should not be used with SZ basis sets, as the error in the sp-part of the basis is much larger than the error from the lack of polarization functions. A DZP basis set is defined by

adding a single d-function to a DZ-type basis set (and a single p-function is added to hydrogen/helium basis sets). Early applications of TZP-type basis sets often added two or three d-functions and one f-function based on performance criteria for a selection of properties and systems.<sup>81</sup> On the basis of energy analysis for atomic and molecular systems, there is now general agreement that the consistent polarization at the TZ level is two d- and one f-function. Numerical evidence suggest that the consistent QZ and 5Z polarization at HF/DFT levels is 4d2f1g and 6d3f2g1h,<sup>43</sup> respectively, whereas it is 3d2f1g and 4d3f2g1h at correlated levels.<sup>36</sup> The difference in the number of d-functions is related to the difference in basis set convergence, as discussed in *Basis Set Convergence of Independent- and Correlated-Electron Methods*.

Four different strategies have been used for assigning the polarization exponents:

1. A distance or overlap criterion with the valence orbitals;
2. Fitting to atomic natural orbitals;
3. Energy minimization of atomic energies;
4. Energy minimization of molecular energies.

The polarization exponents are in the first strategy assigned based on mathematical criteria, such as having a maximum overlap with the valence s- and p-orbitals, or by requiring the same value of the mean distance to the nucleus. The physical argument for this procedure is that the polarization functions at the HF and DFT levels serve to describe the deformation of the valence orbitals, and they should therefore describe the same physical space. Tatewaki and Huzinaga<sup>61</sup> have suggested that d-polarization exponents can be determined by maximizing the orbital overlap with the valence s- and p-orbitals weighted by the number of s- and p-electrons:

$$\Delta = \frac{n_s \langle \phi_s | \chi_d \rangle + n_p \langle \phi_p | \chi_d \rangle}{n_s + n_p}. \quad (9)$$

Roos and Siegbahn<sup>82</sup> have determined polarization exponents by requiring that the expectation value of the distance to the nucleus is identical:

$$r_{\text{mean}} = \langle \chi_d | \mathbf{r} | \chi_d \rangle = \langle \phi_p | \mathbf{r} | \phi_p \rangle. \quad (10)$$

On the basis of these criteria, the exponents should increase with the angular momentum of the polarization function, and a commonly employed ratio is 1.2 between functions differing by one angular momentum.

The second approach has been used by Koga and coworkers,<sup>83</sup> where polarization functions are determined by fitting basis function exponents and contraction coefficients to atomic natural orbitals (ANO), obtained by diagonalizing the density matrix from a correlated calculation using a primitive basis set including polarization functions. The fitting is performed by minimizing a sum of integrated deviations weighted by a factor of  $r^{-2}$ , as shown in Eq. (11):

$$\Delta = \sum_{k=N+1}^{N+N_f} n_k \Delta_k, \quad (11)$$

$$\Delta_k = \int \frac{(\psi_{\text{ANO},k} - \phi_k)^2}{r^2} dv.$$

The sum includes the correlating orbitals  $k = N + 1, \dots, N + N_f$ , i.e., those that are unoccupied at the HF level ( $N$  is the number of occupied orbitals,  $N_f$  is the number of fitting functions, and  $n_k$  is the natural occupation number).

The polarization exponents at HF and DFT levels cannot be determined by energy criteria at the atomic level, as the energy only depends on the occupied orbitals, but they can be determined by energy minimization on molecular systems. For methods including electron correlation, the polarization functions make a nonzero energy contribution already at the atomic level, and polarization exponents can therefore be determined by minimizing the energy of isolated atoms.

The use of atomic correlation energies to assign polarization exponents has been used for several basis sets (Table 5). The polarization functions for the 6-311G(d) basis set have been determined by MP2 optimizations at the atomic level, with 2d and 3d polarization functions obtained by symmetrically splitting the exponents around the optimum value with a constant factor of 4 between the exponents.<sup>14,81</sup> The cc-pVXZ basis sets assign polarization functions by minimizing the atomic CISD energy, using an even-tempered expansion for multiple polarization functions.<sup>36,38</sup> The  $n\text{ZaP}$  basis sets similarly assign polarization functions by MP2 optimization of an even-tempered set of functions.<sup>40</sup>

The fourth strategy assigns polarization exponents by energy minimization on molecular systems, where the optimum values of course will be molecule specific. The polarization exponents for the 6-31G\* basis set were determined by HF energy optimization of a selection of small molecules,<sup>10,84</sup> whereas the pc- $n$  basis sets employ a BLYP energy optimization.<sup>43</sup> Ahlrichs et al.<sup>85</sup> have used minimization of IEPA energies for molecular systems to determine p-, d-,

**TABLE 5** | Basis Set Parameter Optimization

Label	Name	Atomic sp(d)-Exponents		Polarization Exponents		Contraction	
		Method	Target	Method	Target	Method	Coefficients
Pople	STO-3G	Fitted	Atoms	None		Segmented	Fitted
	6-31G	HF-optimized Identical s- and p-valence exponents	Atoms	HF-optimized Averaged	Molecules	Segmented	HF-optimized
	6-311G	MP2-optimized Identical s- and p-valence exponents	Atoms	MP2-optimized	Atoms	Segmented	MP2-optimized
Ahlrichs Jorge	Def2-XZP	HF-optimized	Atoms	From cc-pVXZ		Segmented	HF-optimized
	XZP	HF-optimized for core MP2-optimized for valence	Atoms	MP2-optimized	Atoms	Segmented	HF-optimized
Sapporo Roos		HF-optimized	Atoms	ANO-fitted	Atom	Segmented	HF-optimized
	ANO-XZP	HF-optimized	Atoms	CISD-optimized Even-tempered	Atoms	General	ANO-coefficients
Dunning	cc-pVXZ	HF-optimized	Atoms	CISD-optimized Even-tempered	Atoms	General	HF-coefficients
Petersson	<i>n</i> ZaP	HF L6-optimized for core MP2 even-tempered optimized for valence	Atoms	MP2 optimized Even-tempered	Atoms	General	HF-coefficients
Jensen	pc- <i>n</i>	BLYP-optimized	Atoms	BLYP-optimized Averaged	Molecules	General	BLYP-coefficients

L6 denoted that exponents are optimized using a sixth-order Legendre function, Eq. (7).

Segmented contraction denotes that only a few primitive functions are part of more than one contracted function.

General contraction denotes that a significant number of primitive functions are part of more than one contracted function.

f-polarization, but the Def2 basis sets employ the corresponding cc-pVXZ polarization functions.<sup>21</sup>

The optimum exponent for a single polarization function is for molecular applications a compromise between two opposing trends. Maximizing the orbital overlap between the polarization and valence functions is important for describing the atomic electron correlation, and this leads to a rather strong dependence of the polarization exponent with atomic number. The optimum exponent for describing the change polarization associated with molecular bonding, on the other hand, depends primarily on the bond distance(s), and this leads to a much smaller variation with atomic number.<sup>86</sup> It is often difficult to achieve a good balance between these two effects when only a single polarization function is used. Having two polarization functions with different exponents alleviates the problem, in analogy with the improvements upon going from an SZ to DZ sp-basis set, and this is one of the reasons why results often improve considerably upon going from a DZP- to a TZP-type basis set.

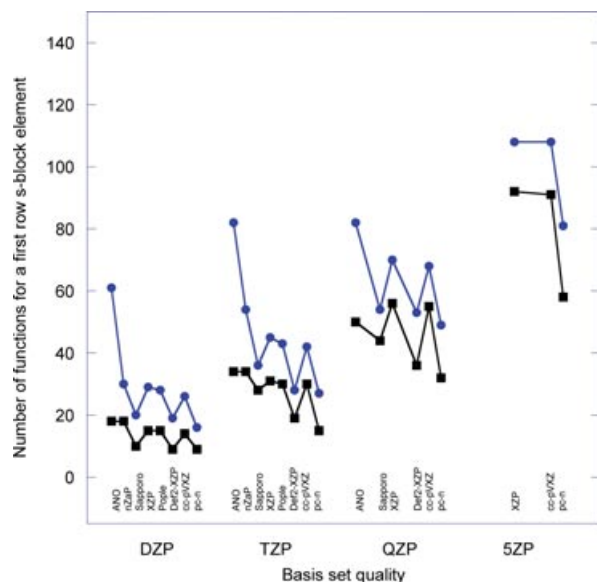
### Core Correlating Basis Sets

The cc-pVXZ basis sets focus, as the ‘V’ implies, on recovering the valence correlation energy. The basis set must be improved by adding correlating functions

with larger exponents if core and core–valence correlation energies are also desired. The cc-pCVXZ basis sets have been designed for this purpose, where the C denotes addition of basis functions with large exponents, often called tight functions. The exponents of the tight functions are assigned based on minimizing the atomic CISD energy, and the number and angular momentum functions to add at each level is again decided based on energy analyses.<sup>87</sup> In subsequent work, the additional tight functions have been optimized with a bias toward recovering intershell correlation, rather than intrashell correlation, as this improves the performance for molecular applications, and these basis sets are denoted cc-pwCVXZ.<sup>39,88</sup> The Sapporo basis sets have also been extended to recover core–valence electron correlation.<sup>89</sup>

### AUGMENTATION WITH DIFFUSE FUNCTIONS

The basis sets described in the previous sections all relate to calculating the energy of neutral systems. If anions are considered, it is clear that the wave function must be improved in regions far from the nuclei, as the extra electron in most cases is relatively weakly bound. It is customary to improve the description of this region of space by adding functions with

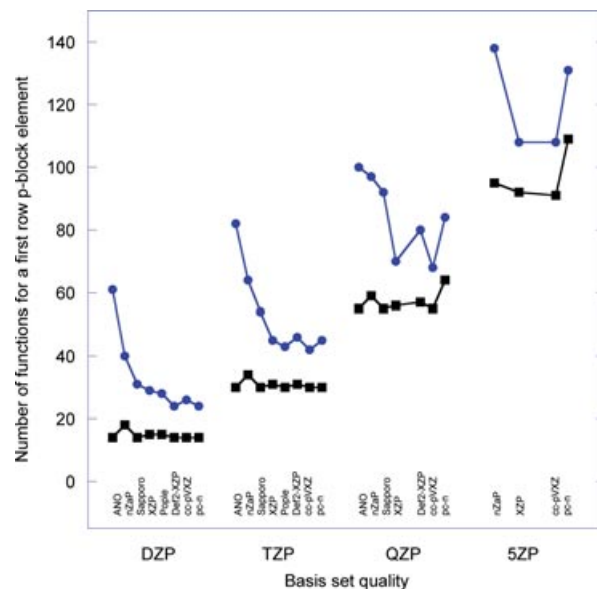


**FIGURE 4** | Number of primitive (blue dots) and contracted (black squares) functions for a first-row s-block atom as a function of basis set quality for the basis sets in Table 1.

small exponents, usually referred to as diffuse functions, as they extend far from the nucleus. Schleyer and coworkers<sup>90</sup> recognized this feature already in 1981, and added a set of diffuse s- and p-type basis functions to the Pople-style *k-lmnG* basis sets. The exponents for the diffuse functions were determined by HF optimization on the atomic anions.<sup>81</sup>

A similar requirement was noted in the development of the cc-pVXZ basis sets, and the exponents for the diffuse functions were in analogy determined by minimizing the energy of atomic anions.<sup>91</sup> In contrast to the Pople-style basis sets, however, a full set of diffuse functions are included up to the highest angular momentum in the basis set. This reflects that the cc-pVXZ basis sets are designed for recovering the correlation energy, which depends on all angular momentum functions, while the Pople-style basis sets focused on the HF energy, where only the occupied atomic orbitals contribute to the energy. Furthermore, many molecular properties require diffuse polarization functions for an adequate description, and the aug-cc-pVXZ basis sets thus allow probing features beyond the energy.<sup>92,93</sup>

Jorge and coworkers<sup>22,94</sup> have similarly defined diffuse augmenting functions by energy optimization of atomic anions, whereas diffuse functions for the Ahlrichs Def2 basis sets have been proposed based on maximizing the atomic polarizability.<sup>95</sup> The *nZaP* basis sets include diffuse s- and p-functions in their definition.



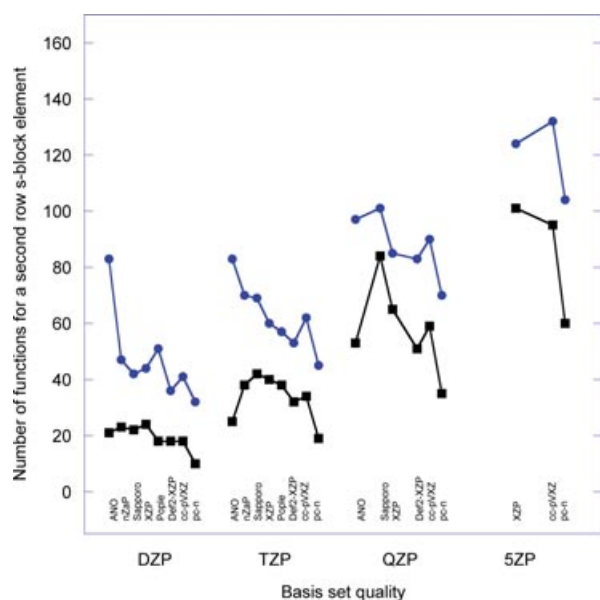
**FIGURE 5** | Number of primitive (blue dots) and contracted (black squares) functions for a first-row p-block atom as a function of basis set quality for the basis sets in Table 1.

The aug-pc-*n* basis sets optimized for DFT methods have similarly been designed to improve the representation of the density in regions far from the nuclei.<sup>96</sup> As most contemporary exchange-correlation functionals have problems describing anions due to self-interaction errors and wrong asymptotic behavior of the potential,<sup>97</sup> it is not possible to define the exponents of the diffuse functions by energy minimization of anions. The exponents of the diffuse functions for the occupied atomic orbitals are instead assigned by scaling the outermost exponents, while the higher angular momentum diffuse functions are determined from an overlap criterion. In analogy with the aug-cc-pVXZ basis sets, the aug-pc-*n* basis sets add diffuse functions for all angular momentum functions included in the basis set to allow a description of a variety of molecular properties.

## CONTRACTION OF BASIS SETS

A primitive basis set is computationally inefficient, but serves as a starting point for constructing contracted basis sets. Tables 1–4 provide an overview of the eight families of basis sets mentioned in the introduction. Figures 4–10 illustrate the compositions for first-, second-, and third-row s-, p-, and d-block elements in terms of contracted and primitive functions for basis sets of DZP to 5ZP quality.



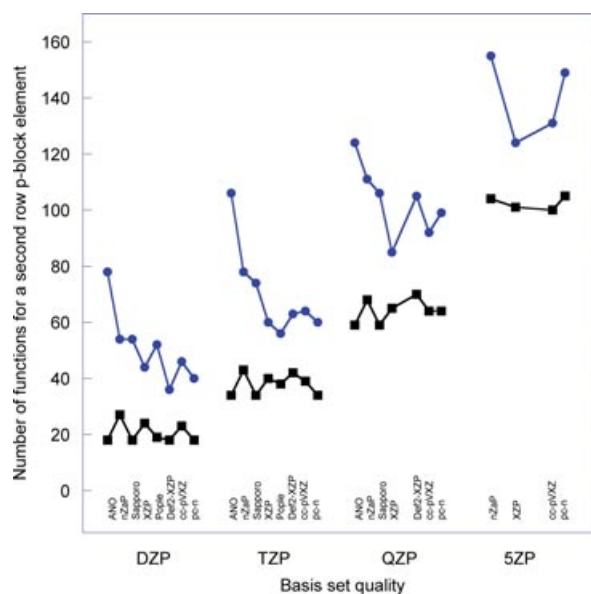


**FIGURE 6** | Number of primitive (blue dots) and contracted (black squares) functions for a second-row s-block atom as a function of basis set quality for the basis sets in Table 2.

### Segmented Contracted Basis Sets

Early work focused on segmented contractions, where the primitive functions were grouped into disjoint partitions based on analyses for small systems.<sup>51,98</sup> The basis set exponents were in this work kept fixed at their optimum values for the uncontracted case, whereas later work employed a simultaneous optimization of both the exponents and contraction coefficients.<sup>63,99</sup>

The most widely known examples of segmented basis sets are the *k-lmnG* basis sets developed by Pople and coworkers. These are of either DZ (*k-lmG*)<sup>8,99–101</sup> or TZ quality (*k-lmnG*),<sup>14</sup> where *k*, *l*, *m*, and *n* indicate the number of primitive functions used in a given contracted function. On the basis of earlier studies with STOs, which showed that there was little difference between the optimum 2s- and 2p-exponents, it was decided to restrict the exponents of the valence s- and p-exponents to be identical, which increased the computational efficiency. With this restriction, all the exponents and contraction coefficients were optimized for atoms at the HF level for 6-31G and at the MP2 level for 6-311G. Note that there are two different versions of the 6-31G basis set for third-row s- and p-block elements (Table 3),<sup>12,13</sup> and different programs employ different defaults. The 6-311G basis set is formally only defined for first-row elements,<sup>14</sup> but similar quality basis sets for second- and third-row s- and p-elements have been published by McLean and Chandler,<sup>15</sup> and by



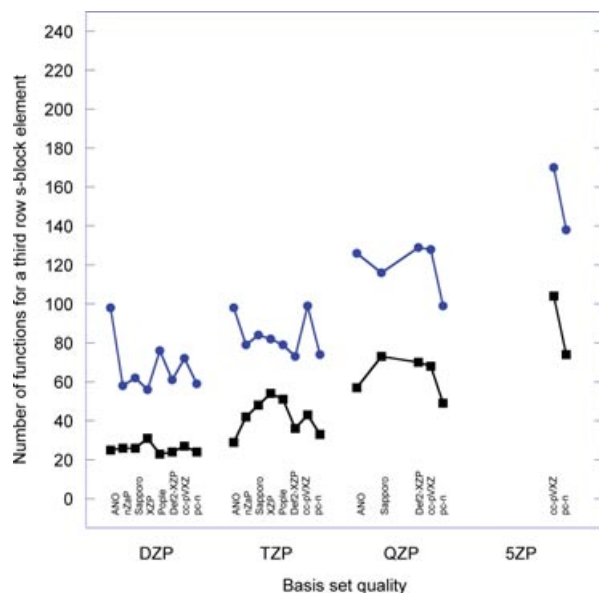
**FIGURE 7** | Number of primitive (blue dots) and contracted (black squares) functions for a second-row p-block atom as a function of basis set quality for the basis sets in Table 2.

Radom and coworkers,<sup>16,17</sup> whereas the Wachters–Hay basis sets are used for the third-row transition metals.<sup>19,20</sup>

Ahlrichs and coworkers<sup>63</sup> have defined SVP (split valence polarized), TZVP,<sup>102</sup> and QZVP<sup>60</sup> basis sets by optimizing both the exponents and contraction coefficients at the HF level for atoms, and these have later been reoptimized to the so-called Def2 versions.<sup>21</sup> Two slightly different sets of polarization functions are available for third-row p- and d-elements (Tables 3 and 4), depending on whether the basis sets are intended used for HF/DFT or correlated methods, where the latter are labeled as TZVPP and QZVPP.

Jorge and coworkers<sup>22,25</sup> have defined XZP basis sets up to 5ZP for first- and second-row elements, and up to TZP for third-row elements, where the atomic core functions are optimized at the HF level, whereas the valence and polarization functions are optimized at the MP2 level. Other groups have reported basis sets optimized at the DFT level, but these are defined only for a limited number of atoms and basis set quality.<sup>103–105</sup>

The Sapporo group has developed DZP, TZP, and QZP quality basis sets where functions describing the occupied atomic orbitals are optimized at the HF level, whereas polarization functions are determined by fitting to CISD atomic natural orbitals.<sup>27–30</sup> These basis sets contain more primitive polarization functions than similar quality basis sets.

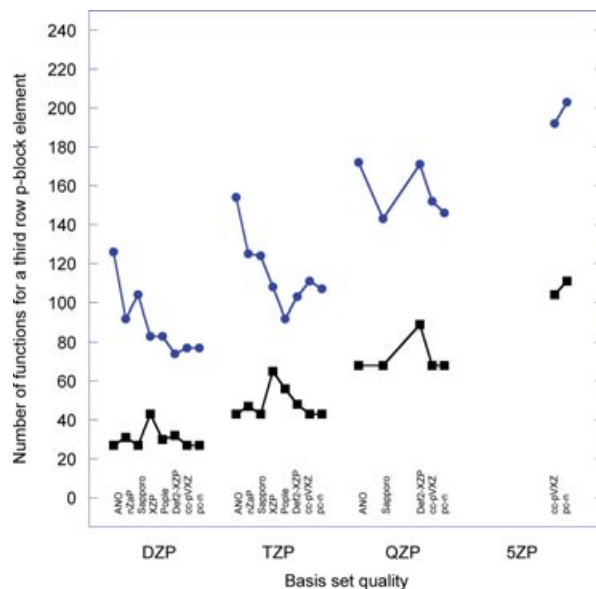


**FIGURE 8** | Number of primitive (blue dots) and contracted (black squares) functions for a third-row s-block atom as a function of basis set quality for the basis sets in Table 3.

One of the main problems with segmented basis sets is the somewhat arbitrary decisions on how many primitive functions to use and how they are contracted to define a given basis set. Even if the number of primitive and contracted functions is decided, there are many possibilities for how the primitives are distributed between the contracted functions, and each of these combinations may furthermore have several local minima in the combined exponent/coefficient parameter space.

Veillard<sup>106</sup> has investigated nine different contractions of a (12s9p) basis set to [6s4p] for second-row p-block atoms. One could argue that the best contraction scheme is the one with the lowest energy, but somewhat discouraging this criterion gives three different contraction schemes for the atoms Si, S, and Ar. The difference is located in the p-contraction, where either a {4212}, {3312}, or {5112} contraction gives the lowest energy. All of these have the outer valence orbital as a contraction of two primitive functions, but chemical intuition suggests that the outer functions should be left uncontracted, i.e., a {5211} or a {6111} contraction will in molecular calculations perform better.

We have similarly shown that there exist at least 19 different minima in the combined exponent and contraction coefficient space for a (11s5p) primitive set of functions contracted to [4s3p] as {6311/311} for the nitrogen atom.<sup>107</sup> The official 6-311G basis set correspond to number seven in terms of energy,



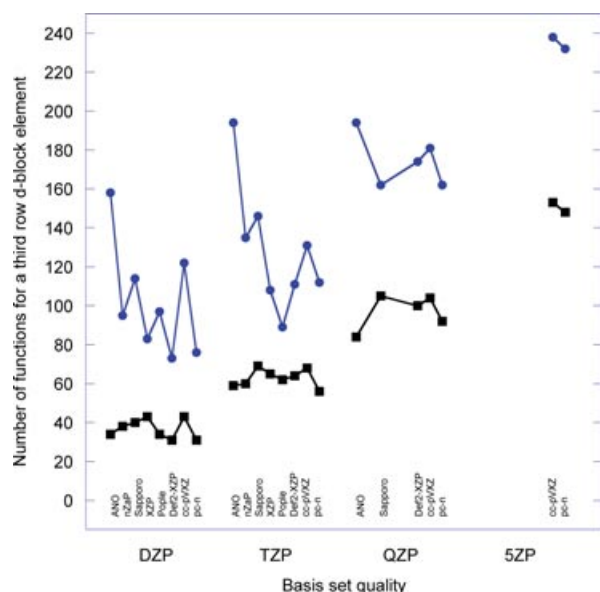
**FIGURE 9** | Number of primitive (blue dots) and contracted (black squares) functions for a third-row p-block atom as a function of basis set quality for the basis sets in Table 3.

but almost all of the 19 exponent/coefficient combinations have very similar performances for molecular calculations. Combining this arbitrariness with the possibility of duplicating one or more primitive functions makes the construction of segmented basis sets somewhat of an art. It should be noted that these problems increase beyond the second row in the periodic table and for constructing basis sets of better than TZ or QZ quality for first and second-row elements. Designing segmented contracted basis sets which systematically approach the basis set limit therefore becomes challenging beyond the TZ level.

## General Contracted Basis Sets

The analysis of natural orbitals suggests that the best approach for designing basis sets for electron correlation is to make a general contraction of a large primitive basis set by using the coefficients from atomic natural orbitals (ANO).<sup>59,108</sup> A single (large) set of primitive functions can be contracted to basis sets of increasingly quality which smoothly converge toward the result for the uncontracted primitive basis set. Roos and coworkers have defined such ANO basis sets, where a common set of primitive functions is contracted to basis sets of DZP, TZP, and QZP quality.<sup>31–34</sup> The main drawback of these basis sets is the computational inefficiency due to the large underlying primitive basis set. The exception is for highly correlated methods, where the cost of integral





**FIGURE 10** | Number of primitive (blue dots) and contracted (black squares) functions for a third-row d-block atom as a function of basis set quality for the basis sets in Table 4.

evaluation is only a minor component of the total computational cost.

Guided by the ideas leading to the ANO concept, Dunning<sup>36</sup> proposed the correlation consistent basis sets (cc-pVXZ) based on a careful analysis of the energy contributions of different basis functions for atoms at the CISD level, which clearly suggested that polarization functions should be included in the order 1d, 2d1f, 3d2f1g, 4d3f2g1h, and so on. The underlying atomic density is described by a general contracted SZ basis set with s- and p-functions, to which there is added one set of correlating s- and p-functions at the DZ level, two sets at the TZ level, and so on. To achieve error balance, the number of primitive functions is also increased for each step up in basis set quality. A SZ basis set is not useful for correlated methods, and the smallest basis set is thus of DZP quality. For second-row elements, it was later discovered that the results could be improved by adding an extra d-function with a larger exponent.<sup>37</sup> These basis sets are sometimes denoted cc-pV(X+d)Z to distinguish them from the original set.

Petersson and coworkers have developed similar types of basis sets, called  $nZaP$  ( $n = 2-6$ ), where the exponents for the core functions are assigned based on HF optimization of a sixth-order Legendre parameterization [Eq. (7)]<sup>62</sup> and valence and polarization functions are taken as a MP2 optimized even-tempered sequence.<sup>40-42</sup> The number of primitive functions is assigned based on an absolute en-

ergy criterion, which leads to a different number of primitive functions for atoms in the same row of the periodic table.

We have developed the polarization consistent basis sets by a general contraction of BLYP optimized primitive functions, where the contraction coefficients are taken as the BLYP atomic orbital coefficients.<sup>43-48</sup> The number of primitive functions is defined by an energy criterion relative to the highest angular momentum function included, and the degree of contraction is determined by requiring that the contraction error is smaller than the inherent error in the primitive basis set, relative to the basis sets limit. Five basis sets of increasing quality have been defined, denoted pc- $n$  ( $n = 0-4$ ), with qualities ranging from (unpolarized) DZ to (polarized) 5ZP.

## BASIS SET CONVERGENCE OF INDEPENDENT- AND CORRELATED-ELECTRON METHODS

The use of basis sets may be considered as a necessary evil, as it introduces a second source of error, besides that inherent in the method used for solving the electronic structure problem. Unraveling the method and basis set errors is often difficult, but is an essential component of method calibration. The use of a hierarchy of basis sets has the significant advantage that it allows an estimate of the (remaining) basis set error, and in favorable cases, permits an extrapolation to the basis set limiting value. A key component in the latter is choosing a suitable mathematical extrapolating function.

Klopper and Kutzelnigg<sup>109-111</sup> showed that the energy of a hydrogen atom converges exponentially with the square-root of the number of Gaussian functions used for expanding the 1s-orbital, as shown in Eq. (12):

$$E_M^{\text{HF}} = E_\infty + Ae^{-B\sqrt{M}}. \quad (12)$$

Figures 2 and 3 show that the HF energy also converges exponentially for many-electron atoms. Close inspection of the curve for the fully optimized exponents shows that it deviates slightly from a straight line. The upward concave shape is qualitatively reproduced by an exponential-square-root expression as in Eq. (12), although the improvement in terms of regression coefficient relative to a straight exponential function is marginal.

Numerical results for molecular systems suggest that the HF energy convergence with respect to angular momentum ( $L$ ) functions also is exponential, but

it is difficult from the numerical data to differentiate between an  $\exp(L)$  or  $\exp(\sqrt{L})$  convergence.<sup>112</sup>

The energy convergence of DFT methods appears to be very similar to HF, which is not surprising, given that the three main components of the energy (electron kinetic energy, and nuclear-electron and electron–electron Coulomb interactions) have identical expressions as in HF theory.<sup>113</sup>

While the basis set at HF or DFT levels only needs to represent the electron density, which is a one-electron quantity, the basis set for the correlation energy also needs to represent the electron–electron correlation cusp, which is a two-electron quantity, and this leads to a difference in convergence behavior.

The convergence of the correlation energy (not the total energy) as a function of angular momentum has been analyzed by Kutzelnigg and Morgan.<sup>114,115</sup> For a basis set saturated up to angular momentum function  $L$ , the remaining error can be expressed as an inverse polynomial in  $L$ , where the leading term is  $L^{-3}$  for a singlet coupled electron pair and  $L^{-5}$  for a triplet coupled pair. This suggests that an inverse polynomial formula such as Eq. (13) may be suitable for extrapolating the correlation energy:

$$E_L^{\text{Correlation}} = E_{\infty} + AL^{-3}. \quad (13)$$

Other formulas including more terms or using an empirical exponent different from 3 have also been used. The underlying physical reason for the significantly slower convergence of the correlation energy compared to HF, is that products of one-electron functions are poor at representing the Coulomb hole surrounding each electron. For a triplet coupled electron pair, the Fermi hole, which is taken into account by the antisymmetric wave function, takes care of the majority of the correlation cusp, and therefore has a faster convergence. For the specific case of MP2, Petersson et al.<sup>116</sup> have shown that the basis set limiting energy can be estimated from a calculation with a single basis set by projecting onto pair natural orbitals.

The difference in convergence behavior of independent- and correlated-electron methods has the consequence that large optimum basis sets for HF/DFT methods have more low-angular momentum functions than basis sets optimized for correlated methods.

The slow basis set convergence of the correlation cusp has motivated the development of methods incorporating a correlation function depending on the interelectronic distance, the so-called F12 methods.<sup>117</sup> These methods have now reached maturity, and as the correlation cusp is taken into account explicitly, the basis set convergence is inverse polynomial with

the leading term being proportional to  $L^{-7}$ . Peterson and coworkers<sup>118,119</sup> have used the same principles as for designing the cc-pVXZ basis sets to optimize basis sets for F12 methods, and these are denoted cc-pVXZ-F12. Relative to cc-pVXZ they contain both more primitive and contracted functions to achieve a proper balance between the HF and correlation energies.

It can be argued from the theoretical analyses that the HF/DFT energy should be extrapolated by an exponential formula, whereas the correlation energy should be extrapolated separately by an inverse polynomial formula. For non-F12 methods, however, the fast(er) exponential convergence of the HF energy component means that the convergence of the total energy follows the slow component, the electron correlation energy, and inverse polynomial extrapolations may therefore be suitable also for the total energy.

Exponential extrapolation formulas, such as Eq. (12), are formally three-point extrapolations, but the exponential  $B$ -parameter can to a good approximation be taken as a universal constant, in which case the extrapolation only requires two points. Functional forms based on a single inverse polynomial such as Eq. (13) are inherently two-point extrapolations.

The analysis showing that the  $L^{-3}$  term dominates for the correlation energy is only strictly valid for a basis set saturated up to angular momentum level  $L$ , which is not fulfilled in practice, and higher-order terms such as  $L^{-4}$  and  $L^{-5}$  also contribute. A variety of exponential, multiexponential and inverse polynomial functional forms have been used for extrapolating the energy calculated with a sequence of basis sets toward the basis set limit, and a recent calibration study for atomization energies showed that none of the proposed formulas stand out as being uniquely best.<sup>120</sup> The common feature is that extrapolations using only DZP and TZP results often give only marginal improvements, as a DZP basis set is too small to represent the correlation cusp with any useful accuracy. The improvement from extrapolation increases as larger and larger basis sets are used, i.e., the relative improvements by extrapolating QZP and 5ZP results are larger than by extrapolating TZP and QZP data. In favorable cases, extrapolation can provide results of typically one basis set higher quality than the largest calculation, e.g., 5ZP quality results from extrapolation of TZP and QZP data.

The success of extrapolation formulas is rooted in the use of hierarchical basis sets, which in a systematic fashion improves the results for each step up in quality. As any reasonable extrapolation formula reduces the basis set error, at no additional

**BOX 1: WEB-BASED BASIS SET LIBRARIES**

A number of Web sites allow downloading of basis sets in an electronic form suitable for use with different programs: The EMSL library includes Pople, Ahlrichs, cc-pVXZ and pc-*n* basis sets: <https://bse.pnl.gov/bse/portal>

The Ahlrichs basis sets: <http://bases.turbo-forum.com/TBL/tbl.html>

The Jorge basis sets: <http://www.cce.ufes.br/qcgv/pub>

The Sapporo basis sets: <http://setani.sci.hokudai.ac.jp/sapporo>

computational cost, there is little reason not to use extrapolations, but it does of course require that calculations are performed with at least two increasingly larger basis sets. One may argue that this should always be done anyway, to ensure that the obtained results are not artifacts of employing a specific (small) basis set.

Basis set extrapolations allow an estimate of the basis set limiting result for a given electronic structure method (HF, DFT, MP<sub>n</sub>, CCSD, ...). To compare with experimental results, it is also necessary to estimate the remaining error in the method itself, and attempt to extrapolate toward the ‘infinite electron correlation’ limit. A brute force approach is infeasible from a computational point of view, and a variety of composite models has been proposed where results from several different methods and basis sets are used to estimate the quantity directly comparable to the experimental value. The *G<sub>n</sub>* methods have recently been reviewed,<sup>121</sup> and this reference also contains references to similar approaches such as the *W<sub>n</sub>*,<sup>122</sup> ccCA,<sup>123</sup> and related<sup>124</sup> methods.

## CONCLUSION

Applications of electronic structure methods have traditionally relied on what J. A. Pople defined as *models*, i.e., selecting a combination of method and basis set, and calibrating the performance by comparing with experimental results. The success of this approach is evident by the literally thousands of papers describing results obtained at the HF, MP2, and B3LYP levels of theory using the 6-31G\* or similar-sized basis sets. As a quality measure for a given application can only be obtained by comparing with previous performances for related systems and properties, these models tend to develop into self-sustaining loops. A researcher looking for a suitable model to use for a given problem cannot fail to notice the large body of previous work using models like B3LYP/6-31G\*,

and therefore often selects this level, thereby further adding to the reference database.

For a sufficiently large database the performance of a specific model can be assessed by statistical methods, but there is always the risk that a given (new) system may be pathological and display errors far outside the calibration data set. Such a failure will not be revealed by a model-type calculation, as there is no way of assessing the errors in the results from the model itself. The model approach furthermore becomes problematic when used for properties where only a few or no calibration data are available.

The model approach also suffers from difficulties in making a direct comparison between the computational and experimental results. A popular molecular property for calibration is the atomization energy, where the calculated results usually are performed using nonrelativistic methods and employing the rigid-rotor harmonic-oscillator model for including finite temperature corrections. The experimental quantity, on the other hand, includes relativistic effects and anharmonic vibrations, and frequently have larger error bars than reported.

The search for ‘magic’ combinations of methods and basis sets that statistically give good results for a (limited) selection of molecules and properties is a popular topic in the literature, and this has become especially pronounced with the proliferation of DFT methods. A significant fraction of these model calibration studies benefits from cancellation between the method and basis set errors, and from errors in the theoretical model compensating for the differences between the theoretically calculated and the experimental measured quantity. DFT methods employing empirical parameters often fit the parameters based on comparison with experimental data using results obtained with a specific basis set, and the DFT parameters in such cases absorb some of the remaining basis set errors.

The trend in electronic structure methods is a move toward a more rigorous control of the errors in the calculated results without reference to external data. This requires a series of calculations with methods and basis sets that can be systematically improved. There are well-defined hierarchies of computational models in wave function theory that are guaranteed to converge toward the exact result, of which coupled cluster methods currently are the most successful. Density functional methods currently do not hold a status as being systematically improvable, but their favorable combination of good accuracy and low computational cost makes them very useful from an application point of view.

Assessing the basis set incompleteness is an important component in establishing error control of the calculated results. This requires calculations with at least two different basis sets, selected such that there is a systematic improvement between the two. The systematic nature not only guarantees a genuine improvement, but also allows extrapolations to be performed, thereby reducing the remaining basis set error without any additional computational effort.

A practical requirement for selecting a basis set is the availability for a variety of elements. It is also important that diffuse and/or tight functions can be added to improve the convergence for sys-

tems/properties where standard basis sets are known to be deficient. There are now a number of hierarchical basis sets that meet these criteria (Box 1), and this has been the focus of the present review. Some of these basis sets have been optimized for specific types of theoretical methods, while others are designed as general purpose basis sets. Although model-type calculations without doubt will continue to be used in thousands of applications also in the future, one can hope that the continuing increase in computational resources will inspire researchers to use hierarchical basis sets to gain insight into and control the inherent errors in the calculated results.

## ACKNOWLEDGMENTS

The author thanks Prof. G. A. Petersson and Prof. R. Lindh for information regarding the *n*ZaP and ANO basis sets. This work was supported by grants from the Danish Center for Scientific Computation and the Danish Natural Science Research Council.

## REFERENCES

1. Slater JC. Atomic shielding constants. *Phys Rev* 1930, 36:0057–0064.
2. Boys SF. Electronic wave functions. 1. A general method of calculation for the stationary states of any molecular system. *Proc R Soc London A: Math Phys Sci* 1950, 200:542–554.
3. Flad HJ, Hackbusch W, Kolb D, Schneider R. Wavelet approximation of correlated wave functions. I. Basics. *J Chem Phys* 2002, 116:9641–9657.
4. Sekino H, Maeda Y, Yanai T, Harrison RJ. Basis set limit Hartree–Fock and density functional theory response property evaluation by multiresolution multi-wavelet basis. *J Chem Phys* 2008, 129:034111.
5. Bischoff FA, Valeev EF. Low-order tensor approximations for electronic wave functions: Hartree–Fock method with guaranteed precision. *J Chem Phys* 2011, 134:104104.
6. Cao X, Dolg M. Pseudopotentials and modelpotentials. *Wiley Interdiscip Rev: Comput Mol Sci* 2011, 1:200–210.
7. Dolg M, Cao X. Relativistic pseudopotentials: their development and scope of applications. *Chem Rev* 2012, 112:403–480.
8. Hehre WJ, Ditchfie R, Pople JA. Self-consistent molecular-orbital methods. 12. Further Extensions of Gaussian-type basis sets for use in molecular-orbital studies of organic-molecules. *J Chem Phys* 1972, 56:2257–2261.
9. Dill JD, Pople JA. Self-consistent molecular-orbital methods. 15. Extended Gaussian-type basis sets for lithium, beryllium, and boron. *J Chem Phys* 1975, 62:2921–2923.
10. Francl MM, Pietro WJ, Hehre WJ, Binkley JS, Gordon MS, Defrees DJ, Pople JA. Self-consistent molecular-orbital methods. 23. A polarization-type basis set for 2nd-row elements. *J Chem Phys* 1982, 77:3654–3665.
11. Rassolov VA, Pople JA, Ratner MA, Windus TL. 6-31G\* basis set for atoms K through Zn. *J Chem Phys* 1998, 109:1223–1229.
12. Rassolov VA, Ratner MA, Pople JA, Redfern PC, Curtiss LA. 6-31G\* basis set for third-row atoms. *J Comput Chem* 2001, 22:976–984.
13. Binning RC, Curtiss LA. Compact contracted basis-sets for 3rd-row atoms—GA-KR. *J Comput Chem* 1990, 11:1206–1216.
14. Krishnan R, Binkley JS, Seeger R, Pople JA. Self-consistent molecular-orbital methods. 20. Basis set for correlated wave-functions. *J Chem Phys* 1980, 72:650–654.
15. McLean AD, Chandler GS. Contracted Gaussian-basis sets for molecular calculations. 1. Second row atoms,  $Z = 11$ –18. *J Chem Phys* 1980, 72:5639–5648.
16. Curtiss LA, McGrath MP, Blaudeau JP, Davis NE, Binning RC, Radom L. Extension of Gaussian-2



- theory to molecules containing 3rd-row atoms GA-KR. *J Chem Phys* 1995, 103:6104–6113.
17. Blaudeau JP, McGrath MP, Curtiss LA, Radom L. Extension of Gaussian-2 (G2) theory to molecules containing third-row atoms K and Ca. *J Chem Phys* 1997, 107:5016–5021.
  18. Dobbs KD, Hehre WJ. Molecular-orbital theory of the properties of inorganic and organometallic compounds. 5. Extended basis-sets for 1st-row transition-metals. *J Comput Chem* 1987, 8:861–879.
  19. Wachters AJ. Gaussian basis set for molecular wave-functions containing third-row atoms. *J Chem Phys* 1970, 52:1033–1036.
  20. Hay PJ. Gaussian basis sets for molecular calculations—representation of 3D orbitals in transition-metal atoms. *J Chem Phys* 1977, 66:4377–4384.
  21. Weigend F, Ahlrichs R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys Chem Chem Phys* 2005, 7:3297–3305.
  22. Neto AC, Muniz EP, Centoducatte R, Jorge FE. Gaussian basis sets for correlated wave functions. Hydrogen, helium, first- and second-row atoms. *Theochem: J Mol Struct* 2005, 718:219–224.
  23. Jorge FE, Sagrillo PS, de Oliveira AR. Gaussian basis sets of 5 zeta valence quality for correlated wave functions. *Chem Phys Lett* 2006, 432:558–563.
  24. Barbieri PL, Fantin PA, Jorge FE. Gaussian basis sets of triple and quadruple zeta valence quality for correlated wave functions. *Mol Phys* 2006, 104:2945–2954.
  25. Machado SF, Camiletti GG, Canal Neto A, Jorge FE, Jorge RS. Gaussian basis set of triple zeta valence quality for the atoms from K to Kr: application in DFT and CCSD(T) calculations of molecular properties. *Mol Phys* 2009, 107:1713–1727.
  26. Camiletti GG, Machado SF, Jorge FE. Gaussian basis set of double zeta quality for atoms K through Kr: application in DFT calculations of molecular properties. *J Comput Chem* 2008, 29:2434–2444.
  27. Thakkar AJ, Koga T, Saito M, Hoffmeyer RE. Double and quadruple zeta-contracted Gaussian-basis sets for hydrogen through neon. *Int J Quantum Chem* 1993:343–354.
  28. Tatewaki H, Koga T, Takashima H. Contracted Gaussian-type basis functions revisited. 2. Atoms Na through Ar. *Theor Chem Acc* 1997, 96:243–247.
  29. Koga T, Tatewaki H, Matsuyama H, Satoh Y. Contracted Gaussian-type basis functions revisited. III. Atoms K through Kr. *Theor Chem Acc* 1999, 102:105–111.
  30. Noro T, Sekiya M, Koga T, Matsuyama H. Valence and correlated basis sets for the first-row transition atoms from Sc to Zn. *Theor Chem Acc* 2000, 104:146–152.
  31. Widmark PO, Malmqvist PA, Roos BO. Density-matrix averaged atomic natural orbital (ANO) basis-sets for correlated molecular wave-functions. 1. First row atoms. *Theor Chim Acta* 1990, 77:291–306.
  32. Widmark PO, Joakim B, Persson, Roos BO. Density-matrix averaged atomic natural orbital (ANO) basis-sets for correlated molecular wave-functions. 2. Second row atoms. *Theor Chim Acta* 1991, 79:419–432.
  33. Pouamerigo R, Merchan M, Nebotgil I, Widmark PO, Roos BO. Density-matrix averaged atomic natural orbital (ANO) basis-sets for correlated molecular wave-functions. 3. First row transition-metal atoms. *Theor Chim Acta* 1995, 92:149–181.
  34. Pierloot K, Dumez B, Widmark PO, Roos BO. Density-matrix averaged atomic natural orbital (ANO) basis-sets for correlated molecular wave-functions. 4. Medium-size basis-sets for the atoms H-Kr. *Theor Chim Acta* 1995, 90:87–114.
  35. Roos BO, Lindh R, Malmqvist PA, Veryazov V, Widmark PO. Main group atoms and dimers studied with a new relativistic ANO basis set. *J Phys Chem A* 2004, 108:2851–2858.
  36. Dunning TH. Gaussian-basis sets for use in correlated molecular calculations. 1. The atoms boron through neon and hydrogen. *J Chem Phys* 1989, 90:1007–1023.
  37. Dunning TH, Peterson KA, Wilson AK. Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. *J Chem Phys* 2001, 114:9244–9253.
  38. Wilson AK, Woon DE, Peterson KA, Dunning TH. Gaussian basis sets for use in correlated molecular calculations. IX. The atoms gallium through krypton. *J Chem Phys* 1999, 110:7667–7676.
  39. Balabanov NB, Peterson KA. Systematically convergent basis sets for transition metals. I. All-electron correlation consistent basis sets for the 3d elements Sc-Zn. *J Chem Phys* 2005, 123:064107.
  40. Zhong S, Barnes EC, Petersson GA. Uniformly convergent n-tuple-zeta augmented polarized (nZaP) basis sets for complete basis set extrapolations. I. Self-consistent field energies. *J Chem Phys* 2008, 129:184116.
  41. Barnes EC, Petersson GA, Feller D, Peterson KA. The CCSD(T) complete basis set limit for Ne revisited. *J Chem Phys* 2008, 129:194115.
  42. Barnes EC, Petersson GA. MP2/CBS atomic and molecular benchmarks for H through Ar. *J Chem Phys* 2010, 132:114111.
  43. Jensen F. Polarization consistent basis sets: principles. *J Chem Phys* 2001, 115:9113–9125.
  44. Jensen F. Polarization consistent basis sets: principles (vol 115, pg 9113, 2001). *J Chem Phys* 2002, 116:3502–3502.

45. Jensen F, Helgaker T. Polarization consistent basis sets. V. The elements Si-Cl. *J Chem Phys* 2004, 121:3463–3470.
46. Jensen F. Polarization consistent basis sets. 4: the elements He, Li, Be, B, Ne, Na, Mg, Al, and Ar. *J Phys Chem A* 2007, 111:11198–11204.
47. Jensen F. Polarization consistent basis sets. VII. The elements K, Ca, Ga, Ge, As, Se, Br, and Kr. *J Chem Phys* 2012, 136:114107.
48. Jensen F. 2012 (submitted).
49. Helgaker T, Jørgensen P, Olsen J. *Molecular Electronic-Structure Theory*. Chichester: John Wiley & Sons; 2000.
50. Jensen F. *Introduction to Computational Chemistry*. Chichester: John Wiley & Sons; 2007.
51. Dunning TH. Gaussian basis functions for use in molecular calculations. 1. Contraction of (9s5p) atomic basis sets for first-row atoms. *J Chem Phys* 1970, 53:2823–2833.
52. Raffenet Rc. General contraction of Gaussian atomic orbitals—core, valence, polarization, and diffuse basis sets—molecular integral evaluation. *J Chem Phys* 1973, 58:4452–4458.
53. Davidson ER. Comment on Dunning's correlation-consistent basis sets. *Chem Phys Lett* 1996, 260:514–518.
54. Almlöf J, Faegri K, Korsell K. Principles for a direct SCF approach to LCAO-MO *ab initio* calculations. *J Comput Chem* 1982, 3:385–399.
55. Haser M, Ahlrichs R. Improvements on the direct SCF method. *J Comput Chem* 1989, 10:104–111.
56. Lambrecht DS, Ochsenfeld C. Multipole-based integral estimates for the rigorous description of distance dependence in two-electron integrals. *J Chem Phys* 2005, 123:184101–184101.
57. Mulliken RS. Criteria for construction of good self-consistent-field molecular orbital wave functions, and significance of LCAO-MO population analysis. *J Chem Phys* 1962, 36:3428–3439.
58. Clementi E, Popkie H. Study of structure of molecular complexes. 1. Energy surface of a water molecule in field of a lithium positive-ion. *J Chem Phys* 1972, 57:1077–1094.
59. Almlöf J, Taylor PR. General contraction of Gaussian-basis sets. 1. Atomic natural orbitals for 1st-row and 2nd-row atoms. *J Chem Phys* 1987, 86:4070–4077.
60. Weigend F, Furche F, Ahlrichs R. Gaussian basis sets of quadruple zeta valence quality for atoms H–Kr. *J Chem Phys* 2003, 119:12753–12762.
61. Tatewaki H, Huzinaga S. A systematic preparation of new contracted Gaussian-type orbital sets. 3. Second-row atoms from Li through Ne. *J Comput Chem* 1980, 1:205–228.
62. Kari RE, Mezey PG, Csizmadia IG. Quality of Gaussian basis sets—direct optimization of orbital exponents by method of conjugate gradients. *J Chem Phys* 1975, 63:581–585.
63. Schafer A, Horn H, Ahlrichs R. Fully optimized contracted Gaussian-basis sets for atoms Li to Kr. *J Chem Phys* 1992, 97:2571–2577.
64. Faegri K, Almlöf J. Energy-optimized GTO basis-sets for LCAO calculations—a gradient approach. *J Comput Chem* 1986, 7:396–405.
65. Mezey PG, Csizmadia IG, Kari RE. Uniform quality Gaussian basis sets. 2. Multiple optima of small Gaussian basis sets for 1st row elements. *J Chem Phys* 1977, 67:2927–2928.
66. Bardo RD, Ruedenberg K. Even-tempered atomic orbitals. 6. Optimal orbital exponents and optimal contractions of Gaussian primitives for hydrogen, carbon, and oxygen in molecules. *J Chem Phys* 1974, 60:918–931.
67. Schmidt MW, Ruedenberg K. Effective convergence to complete orbital bases and to the atomic Hartree–Fock limit through systematic sequences of Gaussian primitives. *J Chem Phys* 1979, 71:3951–3962.
68. Silver DM, Nieuwpoort WC. Universal Atomic Basis Sets. *Chem Phys Lett* 1978, 57:421–422.
69. Moncrieff D, Wilson S. A universal basis set for high-precision molecular electronic structure studies: correlation effects in the ground states of N-2, CO, BF and NO+. *J Phys B-Atomic Mol Opt Phys* 1998, 31:3819–3841.
70. Jorge FE, de Castro EVR. Improved generator coordinate Hartree–Fock method: application to first-row atoms. *Chem Phys Lett* 1999, 302:454–460.
71. Huzinaga S, Klobukowski M. Well-tempered GTF basis-sets for the atoms K through Xe. *Chem Phys Lett* 1985, 120:509–512.
72. Huzinaga S, Klobukowski M, Tatewaki H. The well-tempered GTF basis-sets and their applications in the SCF calculations on N-2, Co, Na-2, and P-2. *Can J Chem—Revue Canadienne De Chimie* 1985, 63:1812–1828.
73. Petersson GA, Zhong SJ, Montgomery JA, Frisch MJ. On the optimization of Gaussian basis sets. *J Chem Phys* 2003, 118:1101–1109.
74. Cohen AJ, Handy NC. Density functional generalized gradient calculations using Slater basis sets. *J Chem Phys* 2002, 117:1470–1478.
75. Van Lenthe E, Baerends EJ. Optimized Slater-type basis sets for the elements 1–118. *J Comput Chem* 2003, 24:1142–1156.
76. Clementi E, Raimondi DL. Atomic screening constants from scf functions. *J Chem Phys* 1963, 38:2686–2689.

77. Huzinaga S. Gaussian-type functions for polyatomic systems. I. *J Chem Phys* 1965, 42:1293–1302.
78. Hehre WJ, Stewart RF, Pople JA. Self-consistent molecular-orbital methods. I. Use of Gaussian expansions of Slater-type atomic orbitals. *J Chem Phys* 1969, 51:2657–2664.
79. Stewart RF. Small Gaussian expansions of atomic orbitals. *J Chem Phys* 1969, 50:2485–2495.
80. Hehre WJ, Ditchfie R, Stewart RF, Pople JA. Self-consistent molecular orbital methods. 4. Use of Gaussian expansions of Slater-type orbitals—extension to second-row molecules. *J Chem Phys* 1970, 52:2769–2773.
81. Frisch MJ, Pople JA, Binkley JS. Self-consistent molecular-orbital methods. 25. Supplementary functions for Gaussian-basis sets. *J Chem Phys* 1984, 80:3265–3269.
82. Roos B, Siegbahn P. Polarization functions for first and second row atoms in Gaussian type MO-SCF calculations. *Theor Chim Acta* 1970, 17:199–208.
83. Noro T, Sekiya M, Koga T. Contracted polarization functions for the atoms helium through neon. *Theor Chem Acc* 1997, 98:25–32.
84. Harihara PC, Pople JA. Influence of polarization functions on molecular-orbital hydrogenation energies. *Theor Chim Acta* 1973, 28:213–222.
85. Ahlrichs R, Driessler F, Lischka H, Staemmler V, Kutzelnigg W. PNO-CI (pair natural orbital configuration interaction) and CEPA-PNO (coupled electron pair approximation with pair natural orbitals) calculations of molecular systems. 2. Molecules BEH<sub>2</sub>, BH<sub>3</sub>, CH<sub>4</sub>, CH<sub>3</sub>, NH<sub>3</sub> (planar and pyramidal), H<sub>2</sub>O, OH+3, HF and NE atom. *J Chem Phys* 1975, 62:1235–1247.
86. Magnusson E. Supplementary d and f functions in molecular wave-functions at large and small internuclear separations. *J Comput Chem* 1993, 14:67–74.
87. Woon DE, Dunning TH. Gaussian-basis sets for use in correlated molecular calculations. 5. Core-valence basis-sets for boron through neon. *J Chem Phys* 1995, 103:4572–4585.
88. Peterson KA, Dunning TH. Accurate correlation consistent basis sets for molecular core-valence correlation effects: the second row atoms Al–Ar, and the first row atoms B–Ne revisited. *J Chem Phys* 2002, 117:10548–10560.
89. Noro T, Sekiya M, Koga T. Segmented contracted basis sets for atoms H through Xe: Sapporo-(DK)-nZP sets ( $n = D, T, Q$ ). *Theor Chem Acc* 2012, 131.
90. Chandrasekhar J, Andrade JG, Schleyer PV. Efficient and accurate calculation of anion proton affinities. *J Am Chem Soc* 1981, 103:5609–5612.
91. Kendall RA, Dunning TH, Harrison RJ. Electron-affinities of the 1st-row atoms revisited—systematic basis-sets and wave-functions. *J Chem Phys* 1992, 96:6796–6806.
92. Woon DE, Dunning TH. Gaussian-basis sets for use in correlated molecular calculations. 4. Calculation of static electrical response properties. *J Chem Phys* 1994, 100:2975–2988.
93. Peterson KA, Dunning TH. The CO molecule: the role of basis set and correlation treatment in the calculation of molecular properties. *Theochem: J Mol Struct* 1997, 400:93–117.
94. Camiletti GG, Canal Neto A, Jorge FE, Machado SF. Augmented Gaussian basis sets of double and triple zeta valence qualities for the atoms K and Sc–Kr: applications in HF, MP2, and DFT calculations of molecular electric properties. *Theochem: J Mol Struct* 2009, 910:122–125.
95. Rappoport D, Furche F. Property-optimized Gaussian basis sets for molecular response calculations. *J Chem Phys* 2010, 133.
96. Jensen F. Polarization consistent basis sets. III. The importance of diffuse functions. *J Chem Phys* 2002, 117:9234–9240.
97. Jensen F. Describing anions by density functional theory: fractional electron affinity. *J Chem Theory Comput* 2010, 6:2726–2735.
98. Dunning TH. Gaussian basis functions for use in molecular calculations. 3. Contraction of (10s6p) atomic basis sets for first-row atoms. *J Chem Phys* 1971, 55:716–723.
99. Ditchfie R, Hehre WJ, Pople JA. Self-consistent molecular-orbital methods. 9. Extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J Chem Phys* 1971, 54:724–728.
100. Gordon MS, Binkley JS, Pople JA, Pietro WJ, Hehre WJ. Self-consistent molecular-orbital methods. 22. Small split-valence basis-sets for 2nd-row elements. *J Am Chem Soc* 1982, 104:2797–2803.
101. Binkley JS, Pople JA, Hehre WJ. Self-consistent molecular-orbital methods. 21. Small split-valence basis-sets for 1st-row elements. *J Am Chem Soc* 1980, 102:939–947.
102. Schafer A, Huber C, Ahlrichs R. Fully optimized contracted Gaussian-basis sets of triple zeta valence quality for atoms Li to Kr. *J Chem Phys* 1994, 100:5829–5835.
103. Godbout N, Salahub DR, Andzelm J, Wimmer E. Optimization of Gaussian-type basis-sets for local spin-density functional calculations. 1. Boron through neon, optimization technique and validation. *Can J Chem—Revue Canadienne De Chimie* 1992, 70:560–571.
104. Chiodo S, Russo N, Sicilia E. Newly developed basis sets for density functional calculations. *J Comput Chem* 2005, 26:175–183.
105. Porezag D, Pederson MR. Optimization of Gaussian basis sets for density-functional calculations. *Phys Rev A* 1999, 60:2840–2847.



106. Veillard A. Gaussian basis set for molecular wavefunctions containing second-row atoms. *Theor Chim Acta* 1968, 12:405–411.
107. Jensen F. Contracted basis sets for density functional calculations: segmented versus general contraction. *J Chem Phys* 2005, 122:074111.
108. Almlöf J, Taylor PR. Atomic natural orbital (ANO) basis-sets for quantum-chemical calculations. *Adv Quantum Chem* 1991, 22:301–373.
109. Klopper W, Kutzelnigg W. Gaussian-basis sets and the nuclear cusp problem. *Theochem: J Mol Struct* 1986, 135:339–356.
110. Kutzelnigg W. Theory of the expansion of wavefunctions in a Gaussian-basis. *Int J Quantum Chem* 1994, 51:447–463.
111. McKemmish LK, Gill PMW. Gaussian expansions of orbitals. *J Chem Theory Comput* 2012.
112. Jensen F. The basis set convergence of the Hartree–Fock energy for H-3(+), Li-2 and N-2. *Theor Chem Acc* 2000, 104:484–490.
113. Christensen KA, Jensen F. The basis set convergence of the density functional energy for H-2. *Chem Phys Lett* 2000, 317:400–403.
114. Kutzelnigg W, Morgan JD. Rates of convergence of the partial-wave expansions of atomic correlation energies. *J Chem Phys* 1992, 96:4484–4508.
115. Kutzelnigg W. Correction. *J Chem Phys* 1992, 97:8821–8821.
116. Petersson GA, Bennett A, Tensfeldt TG, Allaham MA, Shirley WA, Mantzaris J. A complete basis set model chemistry. 1. The total energies of closed-shell atoms and hydrides of the 1st-row elements. *J Chem Phys* 1988, 89:2193–2218.
117. Klopper W, Manby FR, Ten-No S, Valeev EF. R12 methods in explicitly correlated molecular electronic structure theory. *Int Rev Phys Chem* 2006, 25:427–468.
118. Peterson KA, Adler TB, Werner H-J. Systematically convergent basis sets for explicitly correlated wavefunctions: the atoms H, He, B-Ne, and Al-Ar. *J Chem Phys* 2008, 128:084102.
119. Hill JG, Peterson KA. Correlation consistent basis sets for explicitly correlated wavefunctions: valence and core-valence basis sets for Li, Be, Na, and Mg. *Phys Chem Chem Phys* 2010, 12:10460–10468.
120. Feller D, Peterson KA, Hill JG. On the effectiveness of CCSD(T) complete basis set extrapolations for atomization energies. *J Chem Phys* 2011, 135:044102.
121. Curtiss LA, Redfern PC, Raghavachari K. *Gn* theory. *Wiley Interdiscip Rev: Comput Mol Sci* 2011, 1:810–825.
122. Karton A, Rabinovich E, Martin JML, Ruscic B. W4 theory for computational thermochemistry: in pursuit of confident sub-kJ/mol predictions. *J Chem Phys* 2006, 125:144108.
123. DeYonker NJ, Cundari TR, Wilson AK. The correlation consistent composite approach (ccCA): an alternative to the Gaussian-*n* methods. *J Chem Phys* 2006, 124:114104.
124. Feller D, Peterson KA, Dixon DA. A survey of factors contributing to accurate theoretical predictions of atomization energies and molecular structures. *J Chem Phys* 2008, 129:204105.