Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited

Thom H. Dunning, Jr., Kirk A. Peterson, and Angela K. Wilson

Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352

(Received 16 January 2001; accepted 2 March 2001)

For molecules containing second row atoms, unacceptable errors have been found in extrapolating dissociation energies calculated with the standard correlation consistent basis sets to the complete basis set limit. By carefully comparing the convergence behavior of \( D_e (O_2) \) and \( D_e (SO) \), we show that the cause of these errors is a result of two inter-related problems: near duplication of the exponents in two of the \( d \) sets and a lack of high-exponent functions in the early members of the sets. Similar problems exist for the \( f \) sets (and probably in higher angular momentum sets), but have only a minor effect on the calculated dissociation energies. A number of approaches to address the problems in the \( d \) sets were investigated. Well behaved convergence was obtained by augmenting the \((1d)\) and \((2d)\) sets with a high-exponent function and by replacing the \((3d)\) set by the \((4d)\) set and the \((4d)\) set by the \((5d)\) set and so on. To ensure satisfactory coverage of both the \( L \) and \( M \) shell regions, the exponents of the new \( d \) sets were re-optimized. Benchmark calculations on \( Si_2 \), \( PN \), \( SO \), and \( AlCl \) with the new cc-pV\( (n + d)Z \) sets show greatly improved convergence behavior not only for \( D_e \) but for many other properties as well. © 2001 American Institute of Physics.

[DOI: 10.1063/1.1367373]

I. INTRODUCTION

Basis set expansion techniques have been a major contributor to the success of molecular theory by allowing chemists to obtain approximate solutions of the electronic Schrödinger equation for a wide range of molecules. However, truncation of the basis set is also a major source of error in electronic molecular structure calculations. This is a result of the fundamental inability of a basis set expansion in one-electron functions to properly represent the Coulomb cusp, i.e., the behavior of the wave function as the interelectronic distance \((r_{12})\) approaches zero, namely,\(^1\)

\[
\lim_{r_{12} \rightarrow 0} \psi (r_{12}) = 1 + \frac{1}{2} r_{12}.
\]  

As a consequence of this deficiency, molecular properties that are strongly affected by electron correlation converge slowly with increasing basis set size. Because the cost of a calculation depends on a high power of the number of functions \((N)\) in the basis set, e.g., \(N^5\) for MP2 calculations and \(N^7\) for MP4 or CCSD(T) calculations, computational constraints limit the use of large basis sets and, therefore, the accuracy with which many molecular properties can be predicted.

There are two means to address this problem. The first is to explicitly include interelectronic coordinates in the expansion of the wave function to help ensure that Eq. (1) is satisfied. Unfortunately, this approach, although conceptually simple, gives rise to multielectron integrals that are difficult to compute (see, however, the recent paper by Valeev and Schaefer\(^2\)). The R12-method of Kutzelnigg, Klopper, and co-workers,\(^3\) which uses closure relationships to approximate these multielectron integrals, has proven to be very successful in atomic and molecular calculations. Although the use of closure obviates the need to compute multielectron integrals, this approximation only provides accurate solutions to the electronic Schrödinger equation if large, nearly complete basis sets are used—the very problem that we were trying to avoid. The second means to address the problem is to develop a hierarchy of basis sets that systematically approach the complete basis set (CBS) limit, allowing an accurate value of the property to be obtained by extrapolation from results computed with smaller basis sets. This is the approach taken by Dunning and co-workers.\(^4\) This approach is only feasible, of course, if basis sets can be constructed that systematically approach the CBS limit. Fortunately, it appears that the correlation consistent basis sets first introduced by Dunning in 1989,\(^5\) and later extended and elaborated on by Dunning and co-workers,\(^6\) have this property. The sets are denoted as cc-pV\(nZ\), \(n = 2(D), 3(T), 4(Q), 5; \) aug-cc-pV\(nZ\); cc-pCV\(nZ\), etc. Calculations based on this approach have provided some of the most accurate information available on molecular systems, see, e.g., Refs. 4, 7, 8, and 9.

In 1995, Bauschlicher and Partridge\(^10\) reported that they obtained unacceptable errors (~6 kcal/mol) when they extrapolated the binding energy of SO\(_2\) to the CBS limit using the results from CCSD(T) calculations with cc-pVTZ through cc-pV5Z basis sets—the first such failure they had encountered. They found that addition of a high-exponent \(d\)
function to the sulfur basis sets significantly decreased the magnitude of the error in the extrapolation. This issue was later taken up by Martin\textsuperscript{11} who noted that the same problem existed in SO as in SO\textsubscript{2}. He examined the addition of sets of high-exponent functions to the standard correlation consistent basis sets, including the use of \((d,f,g)\) sets that followed the same aufbau pattern used to construct the correlation consistent basis sets. With these additions, significant improvements in the extrapolated binding energies were again found. In the end, Martin recommended the addition of a \((1d)\) set to the cc-pVTZ set, a \((2d1f)\) set to the cc-pVQZ set, and a \((3d2f1g)\) set to the cc-pV5Z set. In an interesting turn of events, Martin showed that the major effect of the tight \(d\) functions was found at the Hartree–Fock level and that it was largely a core polarization effect, rather than a correlation effect. In a related paper, Martin and Uzan\textsuperscript{12} showed that the deficiencies in the correlation consistent basis sets for the second row atoms were not unique to sulfur—other atoms in the second row also exhibited problems indicative of a systematic deficiency in the basis set (Bauschlicher and Partridge\textsuperscript{10} had reported problems with SiO and speculated that the problem might be more general). Shortly after the report by Martin, Bauschlicher and Ricca\textsuperscript{13} examined the effect of various sets of tight functions on the agreement between the CBS limits predicted by different extrapolation procedures. They found the best agreement was achieved with the addition of a \((2d)\) set and recommended the addition of that set to each of the standard cc-pVnZ basis sets for the second row atoms.

In this paper we re-examine the deficiencies in the standard correlation consistent basis sets for the second row atoms.\textsuperscript{6,8} The difficulty in investigating basis set deficiencies is that addition of any function to the basis set will improve the total energy and, since correlation effects are usually larger in the molecule than in the atoms, will likely lead to improvements in the calculated dissociation energy. However, arbitrary additions to a basis set will ruin the systematic behavior needed to provide accurate extrapolations to the complete basis set limit. Thus, one must carefully assess the problem as well as the approach used to solve the problem. By comparing the convergence behavior of \(D_e\) for O\textsubscript{2} and SO, we find two significant, inter-related problems with the \(d\) sets for the second row atoms. Both problems must be addressed to correct the observed convergence problems. As a result, the \(d\) sets put forward here are not the same as those recommended by Martin\textsuperscript{11} or Bauschlicher and Ricca.\textsuperscript{13} For the TZ set, the difference is minor—the exponent of the tight \(d\) function is optimized, rather than estimated. For the QZ-6Z sets, the differences are substantial—the new sets are not obtained by simply augmenting the standard \(d\) sets. Unlike the previous sets, the new sets systematically expand their coverage of the valence and core regions as \(n\) increases. Using the new sets, which we label cc-pV\((n+d)\)Z and aug-cc-pV\((n+d)\)Z, we report calculations on a selection of second row diatomic molecules and compare the results with those obtained with the standard correlation consistent sets. The new sets substantially improve the convergence behavior of \(D_e\) as well as other molecular properties, e.g., \(r_e\) and \(\omega_e\).

### II. COMPUTATIONAL DETAILS

All calculations in this work used the MOLPRO suite of \textit{ab initio} programs.\textsuperscript{14} All optimized exponents (\(\xi\)) in this work were obtained with a BFGS algorithm\textsuperscript{15} using double-sided numerical derivatives. The actual optimizations were carried out in the space of \(\ln(\xi)\), and the gradient of \(\ln(\xi)\) was converged to better than \(1 \times 10^{-6}\). In optimizations involving atomic species, full symmetry equivalencing was used in the orbital calculations. Spectroscopic constants for the diatomics reported in Sec. V were calculated by the usual Dunham analysis\textsuperscript{16} using potential energy functions constructed from standard polynomial fits in internal displacement coordinates to seven total energies (unequally spaced over the range \(-0.3\) bohr \(\leq r - r_e \leq +0.5\) bohr). Dissociation energies were computed with respect to the separated atoms using orbitals optimized with symmetry restrictions consistent with the overall molecular symmetry. In all cases only the pure spherical harmonic components of the polarization functions were used, i.e., \(5d\), \(7f\), etc., and only the valence electrons were correlated.

### III. INCONSISTENCIES AND DEFICIENCIES IN THE STANDARD SULFUR BASIS SETS

Before it is possible to construct basis sets that address any deficiencies in the standard cc-pVnZ sets for the second row atoms, we must better understand and characterize the nature of the defect. Calculated \(E_e\)'s and \(D_e\)'s for O\textsubscript{2} and SO obtained from RCCSD(T)\textsuperscript{17} calculations with the standard correlation consistent sets are listed in Table I. In Fig. 1, we plot the basis set convergence error:

\[
\Delta D_e(n) = D_e(n) - D_e(\infty)
\]

for the two molecules. In Eq. (2), \(D_e(n)\) is the dissociation energy computed with the RCCSD(T) method and a cc-pVnZ basis set, and \(D_e(\infty)\) is the value of the RCCSD(T) dissociation energy at the complete basis set (CBS) limit. \(D_e(\infty)\) was estimated by combining an exponential extrapolation of the Hartree–Fock energies obtained with the cc-pVQZ to cc-pV6Z sets\textsuperscript{18} (see also Ref. 19) with a \(n^{-3}\) extrapolation of the correlation energy obtained with the cc-pVSZ and cc-pV6Z basis sets.\textsuperscript{20} The resulting \(D_e\)'s are 119.98 kcal/mol (O\textsubscript{2}) and 125.64 kcal/mol (SO), which are in good agreement with the corresponding experimental val-

### TABLE I. Total energies \((E_e)\) and dissociation energies \((D_e)\) from RCCSD(T) calculations on O\textsubscript{2} and SO with the correlation consistent basis sets. Total energies in hartrees; dissociation energies in kcal/mol.\textsuperscript{6}

<table>
<thead>
<tr>
<th>Basis set</th>
<th>(E_e)</th>
<th>(D_e)</th>
<th>(E_e)</th>
<th>(D_e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc-pVDZ</td>
<td>-149.985 302 103.980</td>
<td>-472.660 968</td>
<td>94.586</td>
<td></td>
</tr>
<tr>
<td>cc-pVTZ</td>
<td>-150.128 413 113.681</td>
<td>-472.808 798</td>
<td>114.317</td>
<td></td>
</tr>
<tr>
<td>cc-pVQZ</td>
<td>-150.173 221 117.272</td>
<td>-472.851 747</td>
<td>120.428</td>
<td></td>
</tr>
<tr>
<td>cc-pVSZ</td>
<td>-150.188 319 118.570</td>
<td>-472.867 671</td>
<td>123.702</td>
<td></td>
</tr>
<tr>
<td>cc-pV6Z</td>
<td>-150.193 619 119.226</td>
<td>-472.873 286</td>
<td>124.718</td>
<td></td>
</tr>
<tr>
<td>CBS limit</td>
<td>-150.200 35 119.98</td>
<td>-472.879 97</td>
<td>125.64</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{6}Calculated at the fixed experimental bond distances (Ref. 22) of 1.207 52 Å (O\textsubscript{2}) and 1.481 09 Å (SO).
ues (after removing atomic spin–orbit effects\textsuperscript{21}), 119.77 kcal/mol (Ref. 22) (O\textsubscript{2}) and 125.67 kcal/mol (Ref. 23) (SO). As can be seen, the $D^b_{e}(n)$ curve for O\textsubscript{2} smoothly approaches the CBS limit. On the other hand, there is a distinct kink in the $D^b_{e}(n)$ curve, is the cause of the extrapolation errors first noted by Woon and Dunning\textsuperscript{6}.

In Fig. 2, we plot the exponents for the standard $d$ and $f$ sets between $n=4$ (Q) and 5. A discontinuity may also exist in the $g$ sets but is not relevant for the sets being discussed here. The three low-exponent (valence or M-shell) functions in the $(4d)$ and $(4f)$ sets are nearly identical to those in the $(3d)$ and $(3f)$ sets, while the highest exponent functions in the $(4d)$ or $(4f)$ sets are in the core (L-shell) region. Since these sets were optimized in valence electron calculations, Woon and Dunning noted that this was an indication of the need for high-exponent functions to describe correlation effects associated with the electron density in the inner loop of the valence $3p$ orbitals of sulfur. This discontinuity in the $d$-set exponents is the source of the kink in $\Delta D^b_{e}(n)$ observed between the cc-pVQZ and cc-pV5Z sets.

To quantify the contributions to the kink in the $\Delta D^b_{e}(n)$ curve, we carried out a series of calculations on SO to explore the $d$ and $f$ set deficiencies in the L-shell region. In these calculations O\textsubscript{2} was again used as a reference standard. First, we performed calculations with the $d$ sets from the standard cc-pVnZ sets as well as those from the weighted core–valence basis sets, cc-pwCVnZ sets\textsuperscript{24} (note that for oxygen high-exponent $d$ functions are not present in the core–valence sets until the TZ set). The weighted core–valence sets are derived from the standard sets by adding high-exponent functions that systematically expand their coverage of the L-shell region with increasing basis set size. Since the only difference between the two sets is the presence of additional high-exponent $d$ functions in the cc-pwCVnZ sets, the differences between the total energies and the dissociation energies from the two calculations serve to quantify the deficiency in the L-shell region of the standard $d$ set. To minimize problems that might result from deficiencies in the other angular momentum sets, all calculations were performed using the (spfghi)-functions from the cc-pV6Z set.

The results of the above calculations are summarized in Table II and the differences in the calculated $D_e$'s, i.e.,

$$\Delta D^4_{e}(n)=D_e[cc-pV6Z(-d)/cc-pwCVnZ(+d)] - D_e[cc-pV6Z(-d)/cc-pVnZ(+d)]$$

are plotted in Fig. 3. In Eq. (3), $D_e[6Z(-d)/n(+d)]$ refers to calculations with the cc-pV6Z set where the $d$ functions have been replaced by the $d$ functions from the appropriate $nZ$ set. Addition of high-exponent $d$ functions to the oxygen basis sets is seen to have only a minor effect on the calculated $D_e$'s. On the other hand, it is clear that there is a major problem with the standard $d$ sets for the sulfur atom. For the

![FIG. 1. Basis set convergence errors in the dissociation energy, $\Delta D^b_{e}(n)$, from RCCSD(T) calculations on O\textsubscript{2} and SO with the standard cc-pVnZ sets.](image)

![FIG. 2. Exponents for the $d$ and $f$ sets in the standard cc-pVnZ sets of Woon and Dunning [Ref. 6(b)].](image)
TABLE II. Total energies ($E_e$) and dissociation energies ($D_e$) of O$_2$ and SO from RCCSD(T) calculations with the cc-pV6Z sets and the d sets or f sets from either the cc-pV$n$Z sets or the cc-pwCV$n$Z sets. Total energies in hartrees; dissociation energies in kcal/mol.$^a$

<table>
<thead>
<tr>
<th>d/f set source</th>
<th>$E_e$</th>
<th>$D_e$</th>
<th>$E_e$</th>
<th>$D_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc-pVTZ</td>
<td>-150.185 671</td>
<td>119.449</td>
<td>-472.866 568</td>
<td>122.179</td>
</tr>
<tr>
<td>cc-pVQZ</td>
<td>-150.191 628</td>
<td>119.090</td>
<td>-472.869 783</td>
<td>123.156</td>
</tr>
<tr>
<td>cc-pVSZ</td>
<td>-150.193 022</td>
<td>119.140</td>
<td>-472.872 349</td>
<td>124.410</td>
</tr>
<tr>
<td>cc-pwCVTZ</td>
<td>-150.188 526</td>
<td>119.587</td>
<td>-472.872 455</td>
<td>125.046</td>
</tr>
<tr>
<td>cc-pwCVQZ</td>
<td>-150.192 492</td>
<td>119.142</td>
<td>-472.873 295</td>
<td>124.798</td>
</tr>
<tr>
<td>cc-pwCVSZ</td>
<td>-150.193 370</td>
<td>119.149</td>
<td>-472.873 351</td>
<td>124.793</td>
</tr>
</tbody>
</table>

$^a$Calculated at the fixed experimental bond distances (Ref. 22) of 1.207 52 Å (O$_2$) and 1.481 09 Å (SO).

TZ and QZ d sets, $\Delta D_e^{3d}$ for SO is 2.87 and 1.64 kcal/mol, respectively. This is to be compared to 0.14 kcal/mol and 0.05 kcal/mol for O$_2$. For the SZ d set, the difference in SO drops dramatically—to 0.38 kcal/mol. The SZ set, which uses the (4$d$) set, is the first cc-pV$n$Z set that provides a high-exponent d function, see Fig. 2. Clearly, the standard sulfur (1$d$)–(3$d$) sets used in the cc-pVDZ, cc-pVTZ, and cc-pVQZ sets are deficient in high-exponent functions and the differences are chemically significant. The (4$d$) set used in the standard cc-pV5Z set includes a high-exponent d function that begins to address this deficiency. In summary, in agreement with the finding of Bauschlicher and Partridge,$^{10}$ we conclude that the early members of the d sets in the standard cc-pV$n$Z sets for sulfur are deficient in the L-shell region.

Corresponding calculations with the f sets reveal no major differences between O$_2$ and SO as a result of the addition of the high-exponent f functions from the weighted core–valence sets to the standard basis sets (see Table II and Fig. 3). For the QZ set, the first set to contain f functions in O$_2$, $\Delta D_e^{3f}$ for SO is only slightly more than that for O$_2$. This is so in spite of the fact that there are two high-exponent f functions in the sulfur QZ f set versus only one in the oxygen QZ f set. For the SZ f set $\Delta D_e^{3f}$ is about the same for SO as for the QZ set, while $\Delta D_e^{3f}$ for O$_2$ decreases further. However, the magnitude of $\Delta D_e^{5f}$ (SO) is small, 0.031 kcal/mol, as is the SO–O$_2$ differential effect, 0.014 kcal/mol. For the level of accuracy considered here, these differences are not significant. Thus, contrary to the recommendation of Martin,$^{11}$ we do not find it necessary to add high-exponent f functions to the standard cc-pV$n$Z sets for the second row atoms. Further, we do not expect high-exponent functions will be needed by any of the higher angular momentum ($g,h,i$) sets for the current level of accuracy ($\pm$0.1 kcal/mol).

IV. DETERMINATION OF NEW SULFUR BASIS SETS

From the results in the last section, it is clear that we must add high-exponent d functions to the (1$d$)–(3$d$) sets in the standard correlation consistent basis sets. However, the exponents to be used in the final d sets must yield a smooth progression, systematically increasing their coverage of basis function space as $n$ increases. Without this, it will not be possible to extrapolate $D_e^{bn}(n)$ and other molecular properties to the CBS limit. As noted in the last section, the (3$d$) and (4$d$) sets differ only by the addition of a high-exponent function in the (4$d$) set. Were a high-exponent function to be added to each of these sets, as recommended by previous

![FIG. 3. Differences in the calculated dissociation energies for O$_2$ and SO resulting from replacement of the standard d sets with the weighted core–valence d sets ($\Delta D_e^{3d}$) and of the standard f sets with the weighted core–valence f sets ($\Delta D_e^{5f}$).](image-url)
In Table III, we report RCCSD calculations. This effect is present at the Hartree–Fock level. 

We refer to these new sets as the cc-pV(n + d)Z sets. In these sets, each increment in the basis set results in a continuous refinement to the coverage of both the valence (M-shell) and core (L-shell) regions.

We must now determine the best exponents to use in the cc-pV(n + d)Z sets. As noted by Martin, the deficiency of high-exponent \(d\) functions in the standard \(d\)-sets results in a poor representation of core polarization effects in molecular calculations. This effect is present at the Hartree–Fock level.

In Table III, we report RCCSD(T) calculations on SO where we optimized the exponent of the tightest \(d\) function in each of the cc-pV(n + d)Z sets in HF calculations on SO. The optimum exponents, total energies, and dissociation energies are given. Comparing these results with those given in Table I, we see that the \((n + d)\) sets yield dramatic improvements in both the total energy and dissociation energy for the TZ and QZ sets, e.g., the new sets yield \(D_e\)’s for SO that are larger than the standard sets by 2.79 kcal/mol (TZ) and 1.60 kcal/mol (QZ). For the 5Z set, on the other hand, the difference in \(D_e\) is just 0.39 kcal/mol and it decreases to only one-third of that value for the 6Z set. This is just what is required to eliminate the kink in the \(D_e^{\text{bs}}(n)\) curve plotted in Fig. 1.

Although the impact of high-exponent \(d\) functions is largely a molecular effect and dominated by the HF contribution, we felt it worthwhile to determine whether or not reasonable results could be obtained from exponents optimized in atomic calculations. In Table III, the middle columns contain the results of calculations on SO where the exponents of the tightest \(d\) functions in each of the sets were optimized in singles and doubles configuration interaction (CISD) calculations on the ground state of the sulfur atom. The exponents for the TZ and QZ sets are similar to those obtained above as are the total energies and dissociation energies. In fact, the calculated dissociation energies differ by just a few hundredths of a kcal/mol. The optimum exponents for the 5Z and 6Z sets differ more significantly. In spite of this, the resulting differences in total energy and dissociation energy are small. In fact, the total RCCSD(T) energies of SO for these latter two sets are better for the exponents obtained from the atomic CISD calculations than for those from the SO RHF calculations. Thus, atom-optimized basis functions can simultaneously provide an excellent description of both core polarization effects and correlation effects involving the electrons in the inner loops of the sulfur valence orbitals.

To determine if additional \(d\) functions are needed to offset the deficiency in the \(d\) sets, we investigated two different cc-pV(n + 2d)Z sets. The first was obtained by extending the even-tempered series for the \(d\) exponents by two (cc-pVTZ set) or one (cc-pVQZ to cc-pV6Z sets) functions (only one function need be added to the QZ to 6Z sets since these sets already include at least one function in the core region). None of the exponents were re-optimized. The results of these calculations are given in the last columns in Table III. We again include the exponents of the added functions along with the total energies and dissociation energies. As can be seen, the differences between the \((n + d)\) and \((n + 2d)\) sets
TABLE IV. Total energies ($E_n$) and dissociation energies ($D_n$) from RCCSD(T) calculations on $O_2$, SO, and $S_2$ with the optimized aug-cc-pV(n+d)Z sets. Total energies in hartrees; dissociation energies in kcal/mol.

| $n$ | $E_n$ | $D_n$ | $E_n$ | $D_n$ | $E_n$ | $D_n$
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-150.020 305</td>
<td>106.276</td>
<td>-472.711 687</td>
<td>107.864</td>
<td>-795.367 120</td>
<td>86.893</td>
</tr>
<tr>
<td>3</td>
<td>-150.140 319</td>
<td>114.835</td>
<td>-472.826 128</td>
<td>119.593</td>
<td>-795.467 580</td>
<td>96.517</td>
</tr>
<tr>
<td>4</td>
<td>-150.177 984</td>
<td>118.026</td>
<td>-472.859 477</td>
<td>123.182</td>
<td>-795.496 887</td>
<td>100.674</td>
</tr>
<tr>
<td>5</td>
<td>-150.190 316</td>
<td>118.893</td>
<td>-472.870 526</td>
<td>124.505</td>
<td>-795.506 337</td>
<td>102.256</td>
</tr>
<tr>
<td>6</td>
<td>-150.194 591</td>
<td>119.299</td>
<td>-472.874 617</td>
<td>125.023</td>
<td>-795.510 210</td>
<td>102.866</td>
</tr>
<tr>
<td>CBS</td>
<td>-150.200 13</td>
<td>119.84</td>
<td>-472.879 79</td>
<td>125.62</td>
<td>-795.515 23</td>
<td>103.65</td>
</tr>
</tbody>
</table>

*Calculated at the fixed experimental bond distances (Ref. 22) of 1.207 52 Å ($O_2$), 1.481 09 Å (SO), and 1.8892 Å ($S_2$).

Although the $(nd)$ sets determined as described above clearly address the deficiency in the standard $d$ sets, the fact that these sets must span both the $L$- and $M$-shell regions suggests that an even-tempered expansion of the exponents may not be adequate. To ensure that the exponents lead to the optimum progression as $n$ increases, we completely re-optimized the exponents of the $(nd)$ sets in atomic sulfur CISD calculations [the appropriate ($spfg\ldots$) set was used in these optimizations, e.g., the cc-pVQZ ($spfg$h) set was used in optimizing the $(4d)$ set]. This procedure was successful for the $(4d)$--$(6d)$ sets and led to a series of $d$ functions which systematically covered more and more of exponent space (see Fig. 4). However, when the $(3d)$ set was optimized, all of the exponents ended up in the valence space. To determine an appropriate $(3d)$ set, we selected the exponents from the optimum $(2d)$ set, added a high-exponent $d$-function, and optimized its exponent. The resulting $(3d)$ set fits the $(4d)$--$(6d)$ pattern well (see Fig. 4).

The selection of a high-exponent $d$ function for augmenting the standard cc-pVDZ set poses special problems. If one adds a high-exponent $d$ function to the $(1d)$ set and optimizes its exponent, the resulting exponent is much smaller (1.397) than that for the cc-pV($1+2d$)Z set (3.756). This is a result of the poor description of the valence space provided by the single $d$ function in the standard set—the "high" exponent $d$ function is being used to make up for this deficiency. One could argue that the exponent for the $d$ function to be added to the DZ set should be similar to that for the TZ set. Since the low-exponent functions for both of these sets are well localized in the valence region, the function needed to describe core polarization effects should be similar. By analyzing plots of $\xi(nd)$, we decided to estimate the exponent of the tight function for the DZ set by scaling

![Fig. 5. Basis set convergence errors in the dissociation energy, $\Delta D^f_e(n)$, from RCCSD(T) calculations on $O_2$, SO, and $S_2$ with the cc-pV($n+d$)Z and aug-cc-pV($n+d$)Z sets. The $O_2$ calculations used the cc-pVnZ (left figure) and aug-cc-pVnZ (right figure) basis sets; the SO and $S_2$ calculations used the cc-pV($n+d$)Z (left) and aug-cc-pV($n+d$)Z (right) sets.](image-url)
the exponent of the tight function for the TZ set by the ratio \( \zeta_2(TZ)/\zeta_2(QZ) \), yielding 2.994. The resulting (2d) exponents are plotted in Fig. 4. The resulting TZ set is reasonable, although it may yield results slightly better than might be expected for DZ sets in general.

The impact of the L- and M-shell structure on the exponents is clearly evident in Fig. 4. Note, e.g., the large spacing between the highest two exponents in all of the sets and the tighter spacing of \( \zeta_1 \) and \( \zeta_6 \) in the (5d) and (6d) sets. This effect is not well described by an even-tempered series. However, the energetic impact of the effect is not large [<0.2 kcal/mol on \( D_e(SO) \)].

In Table IV we list the total energies and dissociation energies from RCCSD(T) calculations on \( \text{O}_2 \) in the aug-cc-pV\( n \)Z sets and on \( \text{SO} \) and \( \text{S}_2 \) in the new aug-cc-pV\( n+d \) sets. In Fig. 5, we compare the \( \Delta D_{h}^b(n) \) curves for \( \text{O}_2 \) obtained from RCCSD(T) calculations with the cc-pV\( n \)Z and aug-cc-pV\( n \)Z sets to those for \( \text{SO} \) and \( \text{S}_2 \) with the cc-pV\( n+d \)Z and aug-cc-pV\( n+d \)Z sets. For any given basis set \( n \), the magnitude of \( \Delta D_{h}^b(n) \) is smaller for \( \text{O}_2 \) than for \( \text{SO} \) and \( \text{S}_2 \), but the shapes of the curves are now similar. In fact, the \( \text{O}_2 \) and \( \text{S}_2 \) curves are nearly congruent. Addition of diffuse functions to the basis set further reduces the convergence error in \( \text{SO} \), but has a much smaller effect on \( \text{O}_2 \) and \( \text{S}_2 \). This is as expected. SO is more ionic than \( \text{O}_2 \) and \( \text{S}_2 \), and, thus, is more strongly impacted by the addition of diffuse functions.

### V. Benchmark Calculations with \( (n+d) \)-augmented Basis Sets

Upon examining the exponents in the standard \( d \) sets for the second row atoms aluminum through argon, we find that all of the sets exhibit the same discontinuity between \( n_d = 3 \) and \( n_d = 4 \) as in sulfur. The early members of these sets are also deficient in the L-shell region. Thus, the \( d \) sets for all of the second row atoms from aluminum through argon were modified as described above. In Table V we list the \( d \) exponents obtained in this way. These \( d \) sets, when combined with the appropriate \( (spfghi) \) functions from the standard sets, define new cc-pV\( (n+d) \)Z and aug-cc-pV\( (n+d) \)Z basis sets for all of the second row atoms from aluminum through argon.

To determine whether or not the new \( (n+d) \)Z sets correct the convergence problems found in calculations with the standard correlation consistent basis sets, we carried out benchmark calculations on a representative set of molecules.
containing second row atoms: Si$_2$, PN, SO, and AlCl. In Table VI, we report the results of valence electron CCSD(T)/RCCSD(T) calculations on these molecules with the new aug-cc-pV($n+1d$)Z sets. The basis set convergence curves for property ‘Q’, $\Delta Q^b(n)$, are plotted in Figs. 6–8 for $\Delta D_e^b(n)$, $\Delta r_e^b(n)$, and $\Delta \omega_e^b(n)$, respectively. Curves for both the standard aug-cc-pV$n$Z and new aug-cc-pV($n+d$)Z basis sets are given.

The impact of the ($n+d$) sets for the second row atoms is evident in the Figs. 6–8. The $\Delta D_e^b(n)$, $\Delta r_e^b(n)$, and $\Delta \omega_e^b(n)$ curves are far smoother for the aug-cc-pV($n+d$)Z sets than for the aug-cc-pV$n$Z sets. In particular, the pronounced kinks in the $\Delta Q^b(n)$ curves for the aug-cc-pV$n$Z sets between $n=4(Q)$ and 5 are absent in the curves for the aug-cc-pV($n+d$)Z sets. The curves for $\Delta r_e^b(n)$ and $\Delta \omega_e^b(n)$ are less well behaved than $\Delta D_e^b(n)$ with the new sets, but that is almost always found to be the case. However, the $\Delta r_e^b(n)$ and $\Delta \omega_e^b(n)$ curves are better behaved for the new sets than for the standard sets.

In general, the spectroscopic constants predicted by the CCSD(T)/RCCSD(T) calculations agree well with the constants obtained from the experimental data. The difference between $D_e$(CBS) and $D_e$(expt) for PN ($2.9$ kcal/mol) seems unduly large and suggests that the experimental value may be in error. More definitive statements about the accuracy of the coupled cluster method for molecules containing second rows...
row atoms will have to await calculations that include core–valence and relativistic effects.

VI. CONCLUSIONS

For molecules containing the second row atoms, prior authors found unusually large errors in the dissociation energies extrapolated from calculations with the correlation consistent basis sets.\textsuperscript{10–13} Calculations by these authors suggested that the error could be dramatically reduced by adding high exponent functions to the standard correlation consistent basis sets. However, the remedies varied from adding a single extra \( d \) function to the sets,\textsuperscript{10–12} to adding two \( d \) functions to the sets,\textsuperscript{11,13} to adding progressively larger sets of \((d,f,g)\) functions to the sets.\textsuperscript{11}

In this work, the convergence behaviors of \( D_e \) for \( \text{O}_2 \) and \( \text{SO} \) were carefully compared. This comparison shows that the cause of the error in the extrapolated value of \( D_e \) with the standard correlation consistent sets is a result of two inter-related problems in the \( d \) sets: a near duplication of the M-shell exponents in the \((3d)\) and \((4d)\) sets and a deficiency in the L-shell region in the early members of the \( d \) sets needed to describe molecular core polarization effects as well as valence orbital correlation effects. A number of approaches to address these problems were investigated. Satisfactory convergence behavior was obtained by adding one additional high-exponent \( d \) function to the cc-pVDZ and cc-pVTZ sets, and then substituting the \((4d)\) set for the \((3d)\) set in the cc-pVQZ set, the \((5d)\) set for the \((4d)\) set in the cc-pV5Z set, and so on. The exponent for the extra tight \( d \) function in the sulfur TZ set was optimized in correlated atomic calculations as were all of the \( d \) exponents in the remaining \((nd)\) sets \((n=4–6)\). The value of the high-exponent \( d \) function in the DZ set was obtained by scaling the exponent in the TZ set. \( D_e(\text{SO}) \) and \( D_e(\text{S}_2) \) calculated with the new sets, referred to as cc-pV\((n+d)\)Z sets, smoothly converge to the complete basis set limit.

Similar problems have been observed in the basis sets for all of the second row atoms.\textsuperscript{10,12} So, new cc-pV\((n+d)\)Z and aug-cc-pV\((n+d)\)Z sets were generated for all of the second row atoms, aluminum through argon. Benchmark calculations on \( \text{Si}_2 \), \( \text{PN} \), \( \text{SO} \), and \( \text{AlCl} \) with the aug-cc-pV\((n+d)\)Z sets show greatly improved convergence behavior, not only for \( D_e \) but for other properties as well. The new basis sets for the second row atoms, aluminum through argon, are available in the EMSL Gaussian Basis Set Library.\textsuperscript{25}

ACKNOWLEDGMENTS

This work was supported by the Division of Chemical Sciences in the Office of Basic Energy Sciences of the U.S. Department of Energy at Pacific Northwest National Laboratory (PNNL), a multiprogram national laboratory operated by Battelle Memorial Institute, under Contract No. DE-AC06-76RL01830. This research was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) at PNNL. The EMSL is a national user facility funded by the Office of Biological and Environmental Research in the U.S. Department of Energy.


