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

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

Additional information on J. Chem. Phys.

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

Thom H. Dunning, Jr., Kirk A. Peterson,[a)] and Angela K. Wilson[b)]
Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory,[c)] Richland, Washington 99352

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 molecular electronic 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]

$$\text{Lim}_{r_{12}\to 0}\Psi(r_{12})\to 1+\tfrac{1}{2}r_{12}. \tag{1}$$

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 sat-isfied. 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-pVnZ, $n=2$(D), 3(T), 4(Q), 5. ...; aug-cc-pVnZ; cc-pCVnZ, 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

---

[a)] Also at Department of Chemistry, Washington State University, Richland, WA 99352.
[b)] Current address: Department of Chemistry, University of North Texas, Denton, TX 76203.

function to the sulfur basis sets significantly decreased the magnitude of the error in the extrapolation. This issue was later taken up by Martin[11] who noted that the same problem existed in SO as in SO_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[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[10] had reported problems with SiO and speculated that the problem might be more general). Shortly after the report by Martin, Bauschlicher and Ricca[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.[6(b)] 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_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[11] or Bauschlicher and Ricca.[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 ω_e.

TABLE I. Total energies (E_e) and dissociation energies (D_e) from RCCSD(T) calculations on O_2 and SO with the correlation consistent basis sets. Total energies in hartrees; dissociation energies in kcal/mol.[a]

Basis set	O_2		SO	
	E_e	D_e	E_e	D_e
cc-pVDZ	−149.985 302	103.980	−472.660 968	94.586
cc-pVTZ	−150.128 413	113.681	−472.808 798	114.317
cc-pVQZ	−150.173 221	117.272	−472.851 747	120.428
cc-pV5Z	−150.188 319	118.570	−472.867 671	123.702
cc-pV6Z	−150.193 619	119.226	−472.873 286	124.718
CBS limit	*−150.200 35*	*119.98*	*−472.879 97*	*125.64*

[a]Calculated at the fixed experimental bond distances (Ref. 22) of 1.207 52 Å (O_2) and 1.481 09 Å (SO).

II. COMPUTATIONAL DETAILS

All calculations in this work used the MOLPRO suite of *ab initio* programs.[14] All optimized exponents (ζ) in this work were obtained with a BFGS algorithm[15] using double-sided numerical derivatives. The actual optimizations were carried out in the space of $\ln(\zeta)$, and the gradient of $\ln(\zeta)$ was converged to better than 1×10^{-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[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_2 and SO obtained from RCCSD(T)[17] calculations with the standard correlation consistent sets are listed in Table I. In Fig. 1, we plot the basis set convergence error:

$$\Delta D_e^{bs}(n) = D_e(n) - D_e(\infty) \qquad (2)$$

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[18] (see also Ref. 19) with a n^{-3} extrapolation of the correlation energy obtained with the cc-pV5Z and cc-pV6Z basis sets.[20] The resulting D_e's are 119.98 kcal/mol (O_2) and 125.64 kcal/mol (SO), which are in good agreement with the corresponding experimental val-
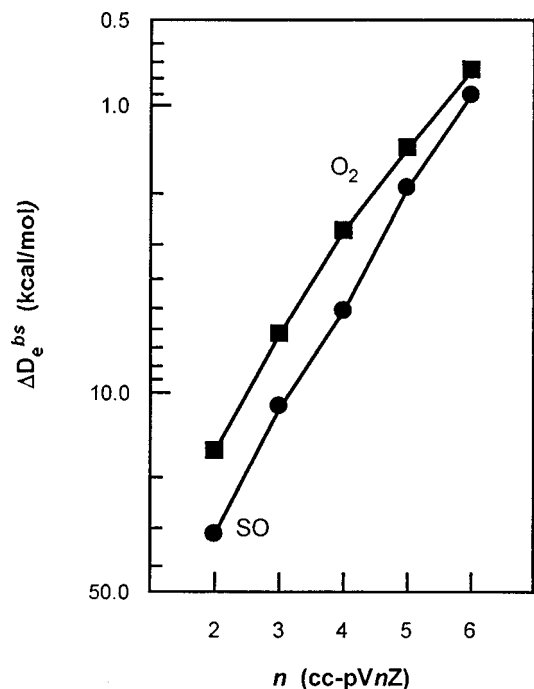
FIG. 1. Basis set convergence errors in the dissociation energy, $\Delta D_e^{bs}(n)$, from RCCSD(T) calculations on O_2 and SO with the standard cc-pVnZ sets.

ues (after removing atomic spin–orbit effects[21]): 119.77 kcal/mol (Ref. 22) (O_2) and 125.67 kcal/mol (Ref. 23) (SO). As can be seen, the $D_e^{bs}(n)$ curve for O_2 smoothly approaches the CBS limit. On the other hand, there is a distinct kink in the $D_e^{bs}(n)$ curve for SO between $n=4$ (Q) and 5. This kink, which noticeably changes the slope of the $D_e^{bs}(n)$ curve, is the cause of the extrapolation errors first noted by Bauschlicher and Partridge.[10]

In Fig. 2, we plot the exponents for the standard d and f sets for sulfur. As noted earlier by Woon and Dunning,[6(b)] there is a clear discontinuity in the exponent patterns for both the d and f sets between n_d, $n_f=3$ and 4. 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 $\Delta D_e^{bs}(n)$ observed between the cc-pVQZ and cc-pV5Z sets.

To quantify the contributions to the kink in the $\Delta D_e^{bs}(n)$ 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_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[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.,

$$\Delta D_e^{\Delta d}(n) = D_e[\text{cc-pV6Z}(-d)/\text{cc-pwCV}n\text{Z}(+d)]$$
$$- D_e[\text{cc-pV6Z}(-d)/\text{cc-pV}n\text{Z}(+d)] \qquad (3)$$

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
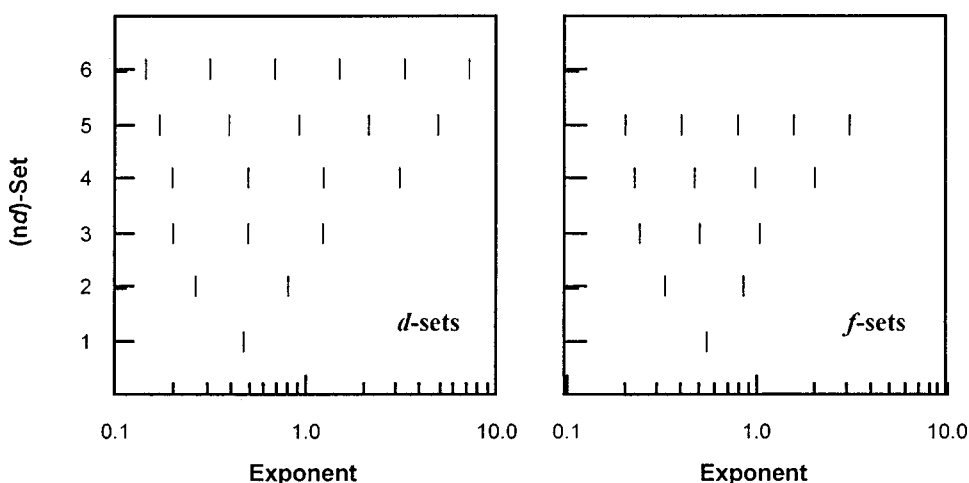

FIG. 2. Exponents for the d and f sets in the standard cc-pVnZ sets of Woon and Dunning [Ref. 6(b)].

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-pVnZ sets or the cc-pwCVnZ sets. Total energies in hartrees; dissociation energies in kcal/mol.[a]

d/f set source	O_2		SO	
	E_e	D_e	E_e	D_e
		d sets		
cc-pVTZ	−150.185 671	119.449	−472.866 568	122.179
cc-pVQZ	−150.191 628	119.090	−472.869 783	123.156
cc-pV5Z	−150.193 022	119.140	−472.872 349	124.410
cc-pwCVTZ	−150.188 526	119.587	−472.872 455	125.046
cc-pwCVQZ	−150.192 492	119.142	−472.873 295	124.798
cc-pwCV5Z	−150.193 370	119.149	−472.873 351	124.793
		f sets		
cc-pVTZ			−472.870 334	125.389
cc-pVQZ	−150.191 163	119.169	−472.872 604	124.676
cc-pV5Z	−150.192 998	119.108	−472.872 992	124.643
cc-pwCVTZ			−472.871 237	125.441
cc-pwCVQZ	−150.191 844	119.194	−472.872 818	124.706
cc-pwCV5Z	−150.193 297	119.125	−472.873 154	124.674

[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, $\Delta D_e^{\Delta d}$ 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 5Z d set, the difference in SO drops dramatically—to 0.38 kcal/mol. The 5Z set, which uses the $(4d)$ set, is the first cc-pVnZ set that provides a high-exponent d function, see Fig. 2. Clearly, the standard sulfur $(1d)$–$(3d)$ 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 $(4d)$ 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-pVnZ 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, $\Delta D_e^{\Delta f}$ 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 5Z f set $\Delta D_e^{\Delta f}$ is about the same for SO as for the QZ set, while $\Delta D_e^{\Delta f}$ for O_2 decreases further. However, the magnitude of $\Delta D_e^{\Delta f}$(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-pVnZ 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 (± 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 $(1d)$–$(3d)$ 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 $D_e^{bs}(n)$ and other molecular properties to the CBS limit. As noted in the last section, the $(3d)$ and $(4d)$ sets differ only by the addition of a high-exponent function in the $(4d)$ set. Were a high-exponent function to be added to each of these sets, as recommended by previous
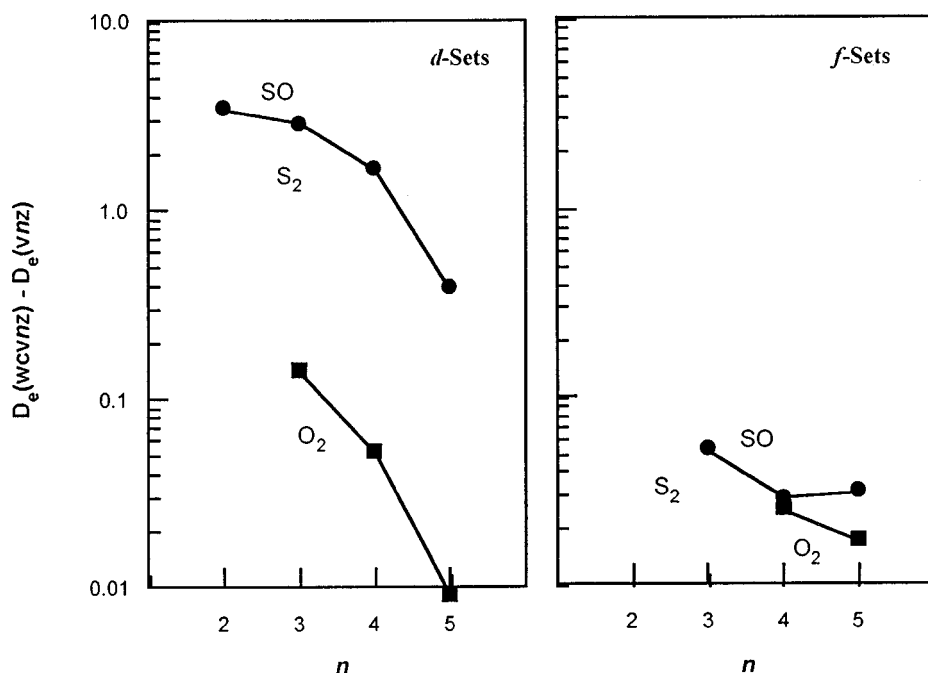

FIG. 3. Differences in the calculated dissociation energies for O_2 and SO resulting from replacement of the standard d sets with the weighted core–valence d sets ($\Delta D_e^{\Delta d}$) and of the standard f sets with the weighted core–valence f sets ($\Delta D_e^{\Delta f}$).

TABLE III. Optimum exponents (ζ), total energies (E_e), and dissociation energies (D_e) of SO from RCCSD(T) calculations with cc-pV($n+d$)Z and cc-pV($n+2d$) basis sets. Total energies are in hartrees; dissociation energies are in kcal/mol.[a]

Basis set	(1d) SO-molecule/HF optimized			(1d) S-atom/CISD optimized			(2d) S-atom/extrapolated		
	ζ	E_e	D_e	ζ	E_e	D_e	ζ	E_e	D_e
cc-pV(T+d)Z	3.564	−472.814 256	117.103	3.987	−472.814 187	117.058	2.494, 7.595	−472.815 007	117.395
cc-pV(Q+d)Z	5.135	−472.854 952	122.033	5.465	−472.854 944	122.031	7.889	−472.855 313	122.146
cc-pV(5+d)Z	10.515	−472.868 562	124.096	6.919	−472.868 644	124.081	11.786	−472.868 792	124.131
cc-pV(6+d)Z	16.706	−472.873 614	124.860	11.711	−472.873 643	124.847	16.019	−472.873 690	124.861
CBS limit		−472.880 01	*125.74*		−472.879 96	*125.73*		−472.879 934	*125.73*

[a]Calculated at the fixed experimental bond distance (Ref. 22) for SO of 1.481 09 Å.

authors,[10–13] the duplication in the valence space of the (3d) and (4d) sets would not be addressed. The magnitude of the problem would be reduced, but it would not be eliminated.

To ensure a smooth progression in the exponents as the d sets are expanded, we use the following prescription.

(i)   Add a high-exponent function to the (1d) and (2d) sets and use them in the DZ and TZ sets, respectively.

(ii) Eliminate the (3d) set, since essentially the same valence functions are present in the (3d) set as in the (4d) set.

(iii) Use the (4d) set for the QZ basis d set, the (5d) set for the 5Z basis d set, and the (6d) set for the 6Z d set.

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,[11] 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 $D_e^{bs}(n)$ 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 off-set 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
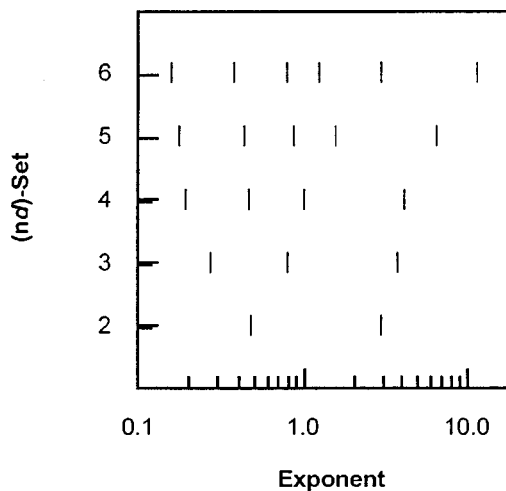

FIG. 4. Exponents for the d sets in the new cc-pV($n+d$)Z sets.

TABLE IV. Total energies (E_e) and dissociation energies (D_e) 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.[a]

	O_2		SO		S_2	
n	E_e	D_e	E_e	D_e	E_e	D_e
2	−150.020 305	106.276	−472.711 687	107.864	−795.367 120	86.893
3	−150.140 319	114.835	−472.826 128	119.593	−795.467 580	96.517
4	−150.177 984	118.026	−472.859 477	123.182	−795.496 887	100.674
5	−150.190 316	118.893	−472.870 526	124.505	−795.506 337	102.256
6	−150.194 591	119.299	−472.874 617	125.023	−795.510 210	102.866
CBS	−150.200 13	119.84	−472.879 79	125.62	−795.515 23	103.65

[a]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).

are small, e.g., the differences in total energy range from 0.82 mE_h (DZ) to 0.05 mE_h (6Z) and in D_e from 0.33 kcal/mol (TZ) to 0.02 kcal/mol (6Z). Thus, contrary to the suggestion of Bauschlicher and Ricca,[13] these results indicate that a $(n+2d)$Z set is not necessary to obtain accurate dissociation energies. To ensure that this conclusion was not biased by our selection of exponents for the $(n+2d)$ augmented sets, we repeated the calculations with the exponents recommended by Bauschlicher and Ricca. The differences between the two $(n+2d)$ sets were negligible, e.g., <0.01 kcal/mol in D_e.

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_d increases, we completely re-optimized the exponents of the (nd) sets in atomic sulfur CISD calculations [the appropriate $(spfg...)$ set was used in these optimizations, e.g., the cc-pVQZ $(spfgh)$ 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(T+d)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 $\zeta_i(nd)$, we decided to estimate the exponent of the tight function for the DZ set by scaling
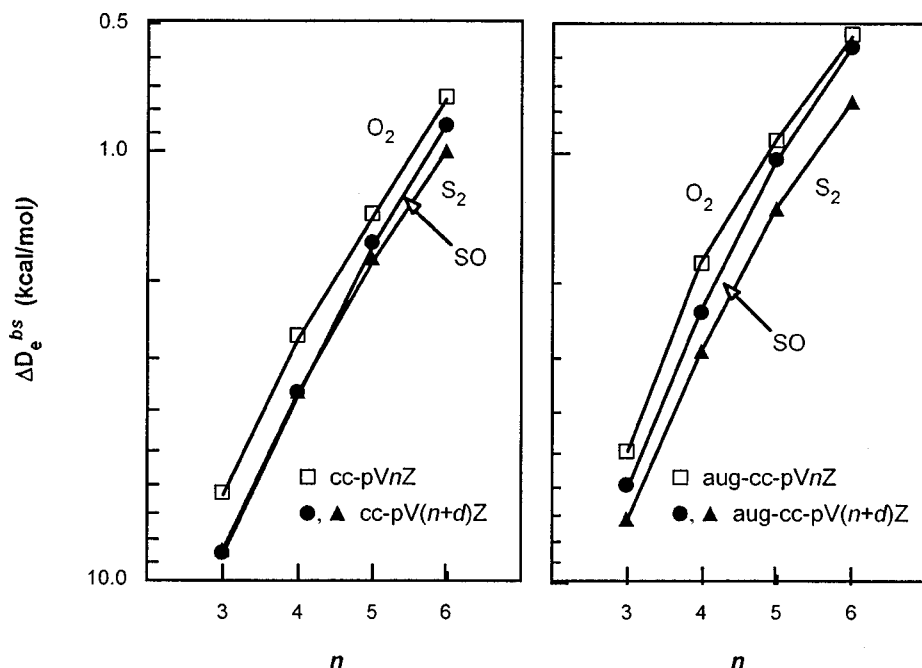

FIG. 5. Basis set convergence errors in the dissociation energy, $\Delta D_e^{bs}(n)$, from RCCSD(T) calculations on O_2, SO, and S_2 with the cc-pV$(n+d)$Z and aug-cc-pV$(n+d)$Z sets. The O_2 calculations used the cc-pVnZ (left figure) and aug-cc-pVnZ (right figure) basis sets; the SO and S_2 calculations used the cc-pV$(n+d)$Z (left) and aug-cc-pV$(n+d)$Z (right) sets.

TABLE V. Exponents in the new d-functions basis sets for the second row atoms, aluminum through argon. The new basis sets are available in the EMSL Gaussian Basis Set Library under the label: cc-pV$(n+d)$Z, where $n=$D, T, Q, 5, and 6.

Atom	Set	Augmented ζ_{aug}	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6
Al	DZ	0.0552	0.190	1.195				
	TZ	0.0357	0.111	0.333	1.570			
	QZ	0.0274	0.0800	0.195	0.437	1.970		
	5Z	0.0254	0.0728	0.180	0.380	0.588	2.639	
	6Z	0.0237	0.0661	0.156	0.350	0.749	0.850	4.240
Si	DZ	0.0840	0.275	1.803				
	TZ	0.0551	0.160	0.476	2.303			
	QZ	0.0414	0.113	0.272	0.608	2.645		
	5Z	0.0390	0.105	0.258	0.533	0.893	3.738	
	6Z	0.0370	0.0977	0.232	0.507	0.835	1.763	6.825
P	DZ	0.116	0.374	2.506				
	TZ	0.0775	0.218	0.648	3.120			
	QZ	0.0570	0.154	0.365	0.807	3.343		
	5Z	0.0537	0.144	0.354	0.724	1.254	5.103	
	6Z	0.0507	0.134	0.319	0.692	1.088	2.482	9.473
S	DZ	0.155	0.481	2.994				
	TZ	0.101	0.273	0.812	3.756			
	QZ	0.0722	0.194	0.464	1.019	4.159		
	5Z	0.0664	0.177	0.438	0.886	1.590	6.510	
	6Z	0.0609	0.160	0.380	0.804	1.259	2.988	11.533
Cl	DZ	0.198	0.603	3.652				
	TZ	0.130	0.339	1.011	4.610			
	QZ	0.0912	0.243	0.583	1.276	5.191		
	5Z	0.0836	0.221	0.549	1.097	2.030	8.399	
	6Z	0.0775	0.203	0.488	1.036	1.583	3.820	14.772
Ar	DZ	0.238	0.739	4.390				
	TZ	0.155	0.412	1.235	5.551			
	QZ	0.108	0.297	0.715	1.562	6.315		
	5Z	0.0978	0.270	0.673	1.332	2.516	10.518	
	6Z	0.0898	0.247	0.594	1.252	1.889	4.580	17.743

TABLE VI. Spectroscopic constants, E_e, D_e, r_e, ω_e, $\omega_e x_e$, and α_e, predicted by CCSD(T) or RCCSD(T) calculations with the new aug-cc-pV$(n+d)$ basis sets on the molecules Si$_2$, PN, SO, and AlCl.

n	E_e (h)	D_e (kcal/mol)	r_e (Å)	ω_e (cm^{-1})	$\omega_e x_e$ (cm^{-1})	α_e (cm^{-1})
			Si$_2$			
2	−577.941 561	64.24	2.2827	497.32	1.97	0.001 31
3	−577.982 314	71.99	2.2624	510.01	1.95	0.001 28
4	−577.994 009	74.47	2.2540	514.29	1.95	0.001 29
5	−577.997 693	75.34	2.2522	515.33	1.93	0.001 29
6	−577.998 988	75.67	2.2514	515.96	1.93	0.001 28
CBS	*−578.000 566*	*76.09*	*2.2501*	*516.78*		
Expt.		75.6	2.2460	511.00	2.02	0.001 30
			PN			
2	−395.482 130	123.32	1.5169	1282.70	6.87	0.005 45
3	−395.557 118	136.98	1.5027	1322.62	6.62	0.005 32
4	−395.579 977	142.27	1.4968	1336.15	6.56	0.005 34
5	−395.587 342	144.06	1.4948	1341.17	6.56	0.005 34
6	−395.590 018	144.79	1.4941	1342.76	6.55	0.005 34
CBS	*−395.593 344*	*145.68*	*1.4933*	*1344.76*		
Expt.		148.6	1.4909	1337.20	6.98	0.005 50
			SO			
2	−472.713 064	108.73	1.5200	1081.37	5.14	0.005 22
3	−472.826 255	119.67	1.4926	1145.22	5.93	0.005 50
4	−472.859 500	123.20	1.4860	1155.30	5.86	0.005 49
5	−472.870 530	124.51	1.4833	1160.30	5.90	0.005 52
6	−472.874 618	125.02	1.4823	1161.94	5.90	0.005 52
CBS	*−472.879 743*	*125.59*	*1.4814*	*1163.89*		
Expt.		126±1	1.4810	1150.80	6.41	0.005 70
			AlCl			
2	−701.720 300	112.35	2.1867	448.58	1.75	0.001 47
3	−701.797 250	118.02	2.1534	473.96	2.01	0.001 53
4	−701.822 171	121.00	2.1433	479.57	2.08	0.001 57
5	−701.829 746	121.90	2.1406	481.17	2.07	0.001 57
6	−701.832 970	122.20	2.1397	481.42	2.08	0.001 58
CBS		*122.58*	*2.1386*	*481.79*		
Expt.		120.9−122.3	2.1301	480.29	2.06	0.001 60

the exponent of the tight function for the TZ set by the ratio ζ_2(TZ)/ζ_3(QZ), yielding 2.994. The resulting $(2d)$ exponents are plotted in Fig. 4. The resulting DZ 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 ζ_3 and ζ_4 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 O$_2$ in the aug-cc-pVnZ sets and on SO and S$_2$ in the new aug-cc-pV$(n+d)$ sets. In Fig. 5, we compare the $\Delta D_e^{bs}(n)$ curves for O$_2$ obtained from RCCSD(T) calculations with the cc-pVnZ and aug-cc-pVnZ sets to those for SO and S$_2$ with the cc-pV$(n+d)$Z and aug-cc-pV$(n+d)$ sets. For any given basis set (n), the magnitude of $\Delta D_e^{bs}(n)$ is smaller for O$_2$ than for SO and S$_2$, but the shapes of the curves are now similar. In fact, the O$_2$ and S$_2$ curves are nearly congruent. Addition of diffuse functions to the basis set further reduces the convergence error in SO, but has a much smaller effect on O$_2$ and S$_2$. This is as expected. SO is more ionic than O$_2$

and 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
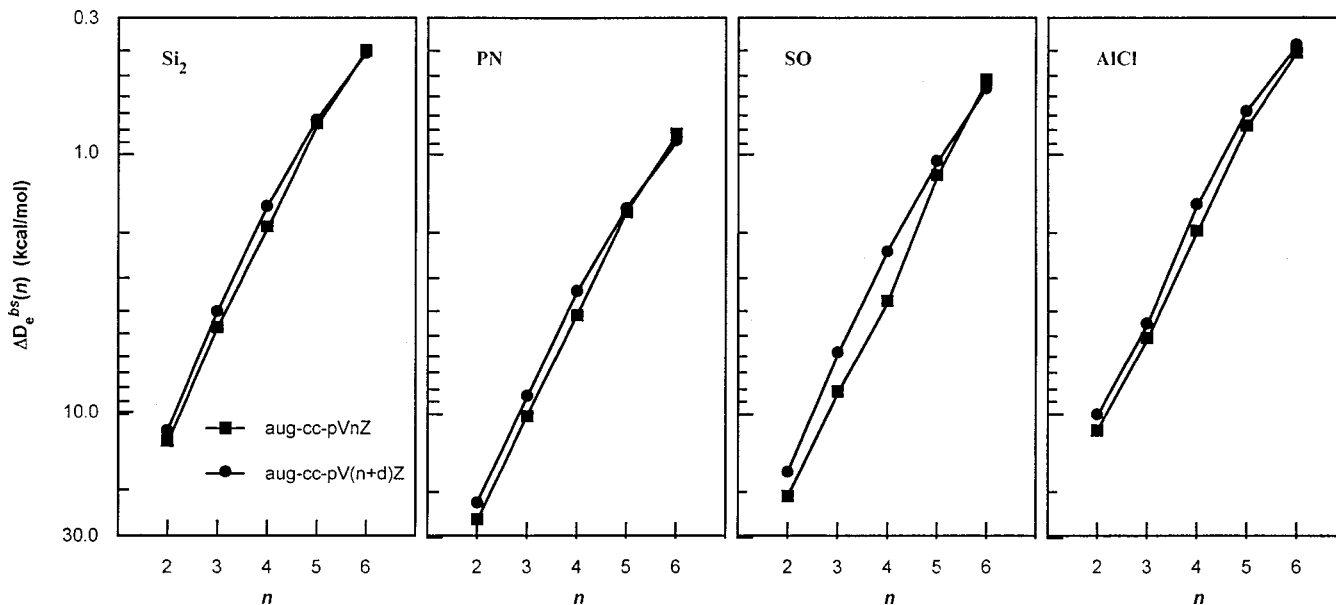
FIG. 6. Basis set convergence errors in the dissociation energy, $\Delta D_e^{bs}(n)$, from CCSD(T)/RCCSD(T) calculations on Si_2, PN, SO, and AlCl with the standard cc-pVnZ and new cc-pV($n+d$)Z sets.

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($n+d$)Z sets. The basis set convergence curves for property "Q", $\Delta Q^{bs}(n)$, are plotted in Figs. 6–8 for $\Delta D_e^{bs}(n)$, $\Delta r_e^{bs}(n)$, and $\Delta \omega_e^{bs}(n)$, respectively. Curves for both the standard aug-cc-pVnZ 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 $\Delta D_e^{bs}(n)$, $\Delta r_e^{bs}(n)$, and $\Delta \omega_e^{bs}(n)$ curves are far smoother for the aug-cc-pV($n+d$)Z sets than for the aug-cc-pVnZ sets. In particular, the pronounced kinks in the $\Delta Q^{bs}(n)$ curves for the

aug-cc-pVnZ sets between $n=4(Q)$ and 5 are absent in the curves for the aug-cc-pV($n+d$)Z sets. The curves for $\Delta r_e^{bs}(n)$ and $\Delta \omega_e^{bs}(n)$ are less well behaved than $\Delta D_e^{bs}(n)$ with the new sets, but that is almost always found to be the case. However, the $\Delta r_e^{bs}(n)$ and $\Delta \omega_e^{bs}(n)$ 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
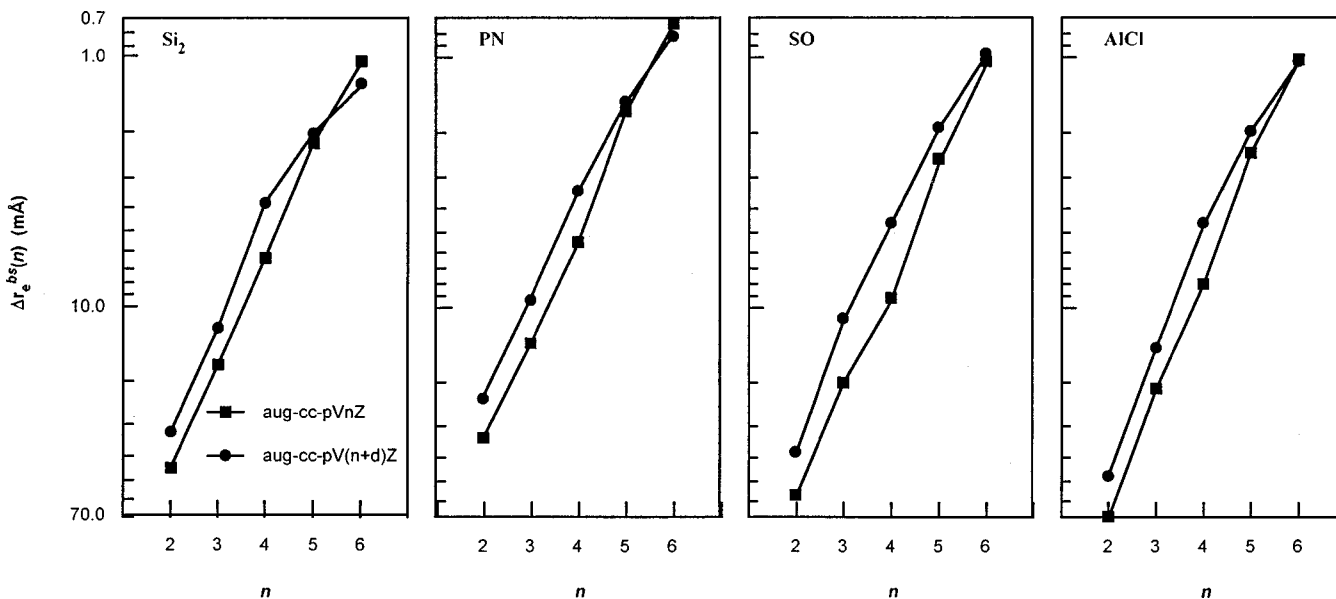

FIG. 7. Basis set convergence errors in the equilibrium bond lengths, $\Delta r_e^{bs}(n)$, from CCSD(T)/RCCSD(T) calculations on Si_2, PN, SO, and AlCl with the standard cc-pVnZ and new cc-pV($n+d$)Z sets.
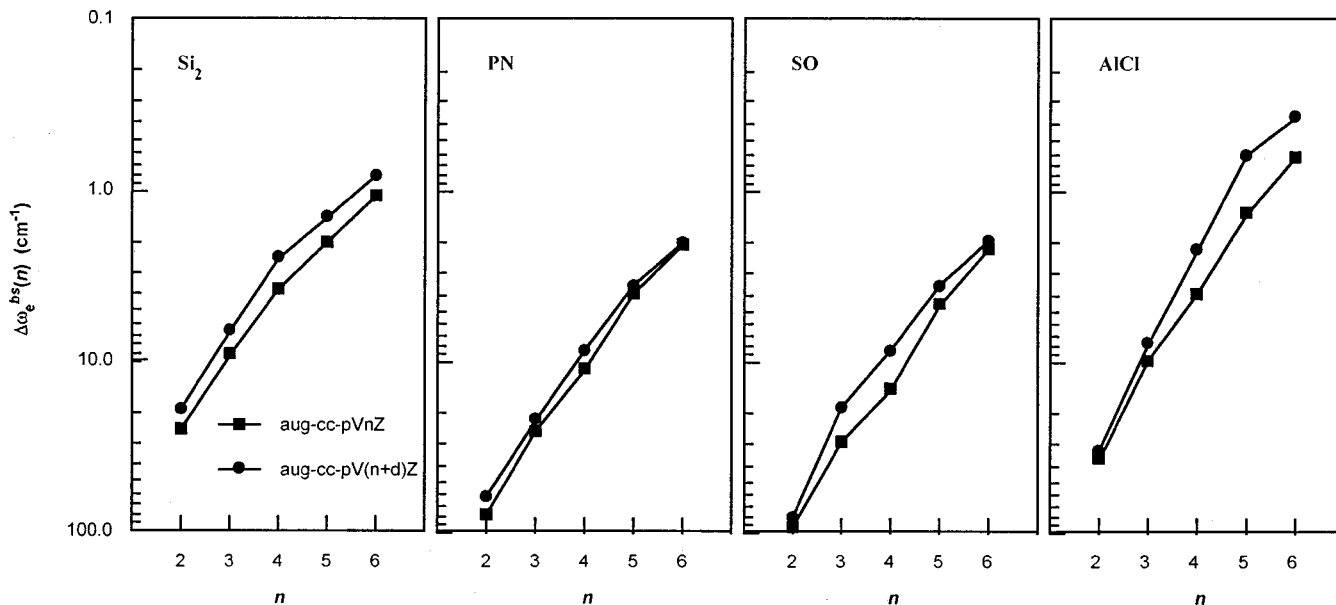
FIG. 8. Basis set convergence errors in the harmonic vibrational frequencies, $\Delta\omega_e^{bs}(n)$, from CCSD(T)/RCCSD(T) calculations on Si_2, PN, SO, and AlCl with the standard cc-pVnZ and new cc-pV($n+d$)Z sets.

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.[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,[10–12] to adding two d functions to the sets,[11,13] to adding progressively larger sets of (d,f,g) functions to the sets.[11]

In this work, the convergence behaviors of D_e for O_2 and 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. $D_e(SO)$ and $D_e(S_2)$ 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.[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 Si_2, PN, SO, and 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.[25]

[1] T. Kato, Commun. Pure Appl. Math. **10**, 151 (1957).

[2] E. F. Valeev and H. F. Schaefer III, J. Chem. Phys. **113**, 3990 (2000).

[3] W. Klopper and W. Kutzelnigg, Chem. Phys. Lett. **134**, 17 (1987); J. Phys. Chem. **94**, 5625 (1990); V. Termath, W. Klopper, and W. Kutzelnigg, J. Chem. Phys. **94**, 2002 (1991); W. Klopper and W. Kutzelnigg, *ibid.* **94**, 2020 (1991); W. Klopper, R. Rohse, and W. Kutzelnigg, Chem. Phys. Lett. **178**, 455 (1991); W. Klopper, *ibid.* **186**, 583 (1991); W. Klopper and J. Almlöf, J. Chem. Phys. **99**, 5167 (1993); W. Klopper, *ibid.* **102**, 6168 (1995); J. Noga, W. Klopper, and W. Kutzelnigg, Chem. Phys. Lett. **199**, 497 (1992); J. Noga and W. Kutzelnigg, J. Chem. Phys. **101**, 7738 (1994); J. Noga, W. Klopper, and W. Kutzelnigg, in *Modern Ideas in Coupled-Cluster Methods*, edited by R. J. Bartlett (World Scientific, Singapore, 1997).

[4] K. A. Peterson and T. H. Dunning, Jr., J. Chem. Phys. **102**, 2032 (1995); J. Phys. Chem. **99**, 3898 (1995); D. E. Woon, T. H. Dunning, Jr., and K.

A. Peterson, J. Chem. Phys. **104**, 5883 (1996); S. S. Xantheas, T. H. Dunning, Jr., and A. Mavridis, *ibid.* **106**, 3280 (1997); K. A. Peterson and T. H. Dunning, Jr., *ibid.* **106**, 4119 (1997); K. A. Peterson and T. H. Dunning, Jr., J. Mol. Struct.: THEOCHEM **400**, 93 (1997); T. van Mourik and T. H. Dunning, Jr., J. Chem. Phys. **107**, 2451 (1997); K. Peterson, A. K. Wilson, D. E. Woon, and T. H. Dunning, Jr., Theor. Chem. Acc. **97**, 251 (1997); D. E. Woon, K. A. Peterson, and T. H. Dunning, Jr., J. Chem. Phys. **109**, 2233 (1998); T. van Mourik, A. K. Wilson, and T. H. Dunning, Jr., Mol. Phys. **96**, 529 (1999).

[5] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).

[6] (a) R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992); (b) D. E. Woon and T. H. Dunning, Jr., *ibid.* **98**, 1358 (1993); (c) D. E. Woon and T. H. Dunning, Jr., *ibid.* **100**, 2975 (1994); (d) D. E. Woon and T. H. Dunning, Jr., *ibid.* **103**, 4572 (1995); (e) A. K. Wilson, T. van Mourik, and T. H. Dunning, Jr., J. Mol. Struct.: THEOCHEM **388**, 339 (1996); (f) A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning, Jr., J. Chem. Phys. **110**, 7667 (1999); (g) T. van Mourik and T. H. Dunning, Jr., Int. J. Quantum Chem. **76**, 205 (2000).

[7] D. Feller, J. Chem. Phys. **96**, 6104 (1992); M. W. Feyereisen, D. Feller, and D. A. Dixon, J. Phys. Chem. **100**, 2993 (1996).

[8] J. M. L. Martin and P. R. Taylor, Chem. Phys. Lett. **248**, 336 (1996); J. M. L. Martin and T. J. Lee, *ibid.* **258**, 129 (1996); **258**, 136 (1996); J. M. L. Martin, *ibid.* **259**, 669 (1996); J. M. L. Martin, *ibid.* **259**, 679 (1996); J. M. L. Martin, in *Computational Thermochemistry. Prediction and Estimation of Molecular Thermodynamics*, edited by K. K. Irikura and D. J. Frurip (American Chemical Society, Washington, DC, 1996); J. M. L. Martin and P. R. Taylor, J. Chem. Phys. **106**, 8620 (1997); J. M. L. Martin, Theor. Chim. Acta **97**, 227 (1997); J. M. L. Martin, Chem. Phys. Lett. **273**, 98 (1997); **292**, 411 (1998); J. M. L. Martin, T. J. Lee, and P. R. Taylor, J. Chem. Phys. **108**, 676 (1998).

[9] A. Halkier, O. Christiansen, D. Sundholm, and P. Pyykkoo, Chem. Phys. Lett. **271**, 273 (1997); A. Halkier, P. Jørgensen, J. Gauss, and T. Helgaker, *ibid.* **274**, 235 (1997); T. Helgaker, J. Gauss, P. Jørgensen, and J. Olsen, J. Chem. Phys. **106**, 6430 (1997); T. Helgaker, W. Klopper, H. Koch, and J. Noga, *ibid.* **106**, 9639 (1997); A. Halkier, H. Koch, P. Jørgensen, O. Christiansen, I. M. Beck Nielsen, and T. Helgaker, Theor. Chim. Acta **97**, 150 (1997); A. Halkier and P. R. Taylor, Chem. Phys. Lett. **285**, 133 (1998); A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, *ibid.* **286**, 243 (1998); A. Halkier, S. Coriani, and P. Jørgensen, *ibid.* **294**, 292 (1998); A. Halkier, H. Larsen, J. Olsen, and P. Jørgensen, J. Chem. Phys. **110**, 734 (1999); A. Halkier and S. Coriani, Chem. Phys. Lett. **303**, 408 (1999); A. Halkier, W. Klopper, T. Helgaker, P. Jørgensen, and P. R. Taylor, J. Chem. Phys. **111**, 9157 (1999).

[10] C. W. Bauschlicher, Jr. and H. Partridge, Chem. Phys. Lett. **240**, 533 (1995).

[11] J. M. L. Martin, J. Chem. Phys. **108**, 2791 (1998).

[12] J. M. L. Martin and O. Uzan, Chem. Phys. Lett. **282**, 16 (1998).

[13] C. W. Bauschlicher, Jr. and A. Ricca, J. Phys. Chem. A **102**, 8044 (1998).

[14] MOLPRO is a package of *ab initio* programs written by H.-J. Werner and P. J. Knowles with contributions from J. Almlöf, R. D. Amos, A. Bernhardsson *et al.*.

[15] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipies in* FORTRAN: the Art of Scientific Computing, 2nd ed. (Cambridge University Press, Cambridge, MA, 1992).

[16] J. L. Dunham, Phys. Rev. **41**, 721 (1932).

[17] G. D. Purvis and R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982); K. Ragavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989); C. Hampel, K. A. Peterson, and H.-J. Werner, *ibid.* **190**, 1 (1992); P. J. Knowles, C. Hampel, and H.-J. Werner, J. Chem. Phys. **99**, 5219 (1994); M. J. O. Deegan and P. J. Knowles, Chem. Phys. Lett. **227**, 321 (1994).

[18] The exponential convergence of the atomic HF energies is illustrated in T. H. Dunning, Jr., K. A. Peterson, and D. E. Woon, in *Encyclopedia of Computational Chemistry,* edited by P. v. R. Schleyer (Wiley, New York, 1998). This estimate will only be reliable when extrapolating the HF energies from the QZ-6Z sets because of the change in slope between the DZ-QZ sets and the QZ-6Z sets.

[19] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).

[20] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).

[21] C. E. Moore, *Atomic Energy Levels* (NSRDS-NBS 35, Office of Standard Reference Data, National Bureau of Standards, Washington, DC, 1971).

[22] K. P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure IV. Constants of Diatomic Molecules* (Van Nostrand, Princeton, NJ, 1979).

[23] G. Clerbaux and R. Colin, J. Mol. Spectrosc. **165**, 334 (1994).

[24] K. A. Peterson and T. H. Dunning, Jr. (unpublished).

[25] D. Feller, http://www.emsl.pnl.gov:2080/forms/basisform.html