

The extrapolation of one-electron basis sets in electronic structure calculations: How it should work and how it can be made to work

David W. Schwenke

MS T27B, NASA Ames Research Center, Moffett Field, California 94035-1000

(Received 16 August 2004; accepted 5 October 2004; published online 14 December 2004)

We consider the extrapolation of the one-electron basis to the basis set limit in the context of coupled cluster calculations. We produce extrapolation coefficients that produce much more accurate results than previous extrapolation forms. These are determined by fitting to accurate benchmark results. For coupled cluster singles doubles energies, we take our benchmark results from the work of Klopper that explicitly includes the interelectronic distance. For the perturbative triples energies, our benchmark results are obtained from large even-tempered basis set calculations.

© 2005 American Institute of Physics. [DOI: 10.1063/1.1824880]

I. INTRODUCTION

In the customary method for solving the electronic Schrödinger equation, one makes a double basis set expansion. The first basis set is the so-called one-electron basis, and these are usually chosen as atomic centered Gaussian radial functions times spherical harmonics. The second basis consists of Slater determinants of different occupations of molecular orbitals built from the one-electron basis functions, or linear combinations of Slater determinants having a particular spin and symmetry eigenvalue (configuration state functions, or CSFs). This second basis accounts for electron correlation. The topic of the present work is accelerating the convergence of the one-electron basis to the complete basis set limit, while electron correlation will be treated using the coupled cluster method including single and double excitation amplitudes (CCSD).¹

A big advance in the field was the introduction of one-electron basis sets derived from atomic natural orbitals (ANO) by Amlöf and Taylor.² These basis sets introduced a scheme whereby the completeness of the radial and angular parts were equally balanced in correlated calculations, and furthermore could be systematically improved. These ideas were developed further by Dunning,³ who developed families of compact basis sets with the same philosophy. The Dunning basis sets have become the defacto standard in electronic structure calculations. The most common family of Dunning basis sets are denoted cc-pVXZ (correlation consistent polarized valence X tuple zeta) with $X=D, T, Q, 5, 6$.

More recently there has been much interest in exploiting the systematic behavior of these basis sets with respect to increasing X by carrying out calculations for several values of X and extrapolating to the basis set limit. This requires knowledge about how the energies scale with X , and there has been considerable debate in the literature on what form is most appropriate. Analytic work by Schwartz⁴ on He and by Kutzelnigg and Morgan⁵ on multielectron atoms using the MP2 (second order Møller–Plesset perturbation theory) method assumes that the one-electron basis is complete for a given l , and yields the asymptotic dependence of the correlation energy on the maximum value of l , which we call l_{\max} .

Since for the first row elements $X=l_{\max}$, it is natural to use X in the asymptotic formulas, yielding

$$E^c(X) = E^c(\infty) + C^c X^{-n}, \quad (1)$$

where c denotes the particular contribution to the energy, with $n=3$ for singlet He or singlet pairs and $n=5$ for triplet pairs. This form arises from consideration of the limiting behavior of the wave function as two electrons coalesce.

It should be carefully noted that c is one of HF (the Hartree-Fock energy), SP (the singlet pair) contribution to the CCSD correlation energy, TP (the triplet pair) contribution to the CCSD correlation energy, or (T) (the perturbative triples corrections). Equation (1) has been used for $c=SP$, $c=TP$, and $c=SP+TP$.

Motivated by the fact that $l_{\max}=X-1$ for hydrogen, Martin⁶ suggested one use,

$$E^c(X) = E^c(\infty) + C^c/(X+a)^3, \quad (2)$$

for the total correlation energy, i.e., $c=SP+TP+(T)$. Peterson⁷ has suggested one use,

$$E^c(X) = E^c(\infty) + \sum_n C_n^c \exp[-(X-1)^n], \quad (3)$$

for the total energy, i.e., $c=HF+SP+TP+(T)$. The formula is motivated by observations that for small basis sets, Eq. (1) yields too large a correction. Initial work⁸ also utilized the expression

$$E^c(X) = E^c(\infty) + A^c \exp(-bX), \quad (4)$$

for the total energy, i.e., $c=HF+SP+TP+(T)$, although it is now recognized that this is more appropriate for the HF energy.⁹

The work of Klopper¹⁰ provides an excellent review of previous work as well as benchmark results obtained using the CCSD method with explicit inclusion of the interelectronic distance (CCSD-R12 method). He carried out large basis set CCSD-R12 calculations and conventional CCSD calculations using the cc-pVXZ basis sets for a test set of the seven species Ne, N₂, ¹A₁ CH₂, H₂O, CO, HF, and F₂, all at their minimum energy geometry determined using the cc-

TABLE I. Geometrical parameters for the molecules.

	R (Å)	\angle (deg)
CH ₂ (¹ A ₁)	1.106 758	102.027
H ₂ O	0.957 119	104.225
HF	0.915 769	
N ₂	1.098 119	
CO	1.128 876	
F ₂	1.411 336	

pCVQZ basis and the CCSD(T) method correlating all electrons. The geometrical parameters for the minima are given in Table I. The CCSD-R12 results are expected to be converged with respect to the one-electron basis, providing accurate results for the seven molecules. Thus this work provides unambiguous results for judging the quality of basis set extrapolation.

Klopper also shows the superiority of decomposing the correlation energy into the singlet pair contribution and triplet pair contribution, and then extrapolating the singlet pair contribution using the scaling X^{-3} and extrapolating the triplet pair contribution using the scaling X^{-5} . He also reports benchmark results of calculations of core-valence energies using the CCSD-R12 method as well as conventional CCSD calculations using cc-pCVXZ basis sets,¹¹ but he does not consider extrapolation of the later results. He also reports (T) energies, but one must bear in mind that these do not explicitly include the interelectronic distance, and thus do not provide accurate benchmark results.

The work of Klopper¹⁰ leaves several unanswered questions. It is well known that molecules such as HF have a great degree of ionic character, and hence require extra diffuse basis functions; thus the aug-cc-pVXZ basis sets¹² are more appropriate. This is also the case when one is looking at weakly bound systems. One can ask whether or not the extrapolation obviates the need for aug-cc-pVXZ basis sets for energies, or the need for counterpoise corrections to the basis set superposition error. Another issue not addressed by Klopper is the convergence of the perturbative triples energies in CCSD(T) calculations. One can also ask whether or not the extrapolation can be further improved, for example, to obtain higher accuracy given a pair of basis sets by looking at departures from Eq. (1) due to the substitution $X = l_{\max}$, or to obtain even higher accuracy when using more than two basis sets by including more terms in the extrapolation formulas. These are questions we will answer in the present work.

II. HOW EXTRAPOLATION SHOULD WORK

The analytic work^{4,5} assumes that the one-electron basis is converged for each value of l . In this section we test how extrapolation works when basis sets of this type are used.

Basis sets for the atoms H, C, N, O, F, and Ne that are converged for each l are not available in the literature, so it was necessary to generate our own. We represented the basis functions for each l as even-tempered¹³ expansions; thus we need to optimize two parameters for each l and each choice of number of basis functions. We performed this optimiza-

tion by means of Powell's method¹⁴ implemented in Molpro¹⁵ by means of Molpro's job control language. For Ne we used the CCSD method, while for the other atoms we used the CISD (configuration interaction singles and doubles) method using orbitals obtained from state averaged MCSCF (multiconfiguration self-consistent field) calculations designed to yield spherical atoms. For H, we optimized basis functions for the H₂ molecule at $r = 1.4a_0$. In all cases we only considered the ground electronic state and only correlated the valence electrons.

When optimizing s functions, we used the p and d (only p for H₂) functions from the cc-pV6Z basis set, and increased the number of functions in steps of 2 until the optimized energies differed by less than about $5\mu E_h$. This largest s set was used in all further optimizations. We then optimized the p, d, \dots shells using the same procedure. For H₂, we optimized up to h functions. For the first row atoms, we observed that the spacing between functions and the center of the functions tended to approach limits as l was increased, while at the same time the number of required functions decreased in a regular manner. We thus only optimized up to g functions and used the spacing and center from the g functions for the h and i functions, while decreasing the number of functions in a regular manner. For Ne, N, C, and O, to be conservative, we used more h and i functions than the optimization trend indicated, while for F we did not use any extra functions. This leads to the following basis sets for H: $15s8p6d6f4g4h$, for C: $23s15p10d8f6g6h4i$, for N: $23s17p12d9f7g7h5i$, for O: $23s15p12d10f8g8h6i$, for F: $23s17p14d12f10g8h6i$, and for Ne: $23s15p14d11f9g9h7i$. We call these f -limit basis sets. We give the parameters for the optimized basis sets in Table II.

We use these basis sets uncontracted in all calculations. Possible economies could be realized by generating ANO's, but we have not done so in the present work. All calculations we run using a modified version of MOLPRO 2000.1 on a SGI Origin 3000 in single processor mode. The two-electron integrals were computed once and stored on disk. These basis sets proved to be too large to run the largest basis for H₂O and F₂.

In Table III we give the SCF energies obtained from the f -limit basis sets. We obtain exponential convergence with respect to increasing l_{\max} , and our best results are in very good agreement with the results of Klopper.

In Table IV we give the singlet and triplet pair contributions to the CCSD correlation energies, and in Table V we give (T) energies for the seven test species computed using the f -limit basis sets.

In Fig. 1, we show the root-mean-square errors in the total CCSD correlation energy over the seven test species for the unextrapolated cc-pVXZ basis sets, the f -limit basis sets, and the results from extrapolating those basis sets using Eq. (1) with $n=3$ for the singlet pair energies and $n=5$ for the triplet pair energies. Consider first the unextrapolated results. As expected, the f -limit basis results have smaller errors, but the two curves run pretty much parallel, with the f -limit basis errors a little more than a factor of two smaller. That the curves run parallel is a tribute to the skill of Dunning in optimizing the basis sets. In contrast, once extrapolation is

TABLE II. Basis function parameters for f -limit basis sets. $\zeta = \alpha\beta^n$, $n = 1, \dots, N_f$.

l	H	C	N	O	F	Ne
	15 ^a	23	23	23	23	23
s	2.102 465 58 ^b 7.486 027 91 ^c 8	2.012 393 44 165.073 510 76 15	2.009 993 70 212.782 754 62 17	2.039 542 08 179.540 982 11 15	2.007 185 77 376.475 468 86 17	2.021 688 64 415.996 424 14 15
p	1.797 386 89 1.304 477 02 6	1.875 144 35 3.143 095 77 10	1.804 640 23 6.945 407 86 12	2.000 000 00 3.275 534 35 12	1.805 564 62 7.752 815 83 14	1.983 261 19 14.537 657 71 14
d	1.858 867 59 1.633 852 49 6	1.760 194 83 1.435 986 98 8	1.720 236 66 3.075 896 41 9	1.664 833 00 2.803 992 39 10	1.548 992 99 3.738 243 56 12	1.559 241 01 4.850 572 08 11
f	1.872 460 35 2.555 765 17 4	1.732 435 75 1.531 308 11 6	1.696 211 04 2.442 677 58 7	1.656 390 86 2.964 889 85 8	1.548 992 99 3.983 487 05 10	1.583 140 72 4.335 992 13 9
g	1.927 388 56 2.628 495 06	1.723 504 91 1.520 327 00	1.710 936 35 2.452 559 58	1.649 757 72 3.107 983 02	1.548 992 99 4.251 004 98	1.583 000 00 4.682 841 62
h	4 1.927 388 56 2.628 495 06	6 1.723 504 91 1.520 327 00 4	7 1.710 936 35 2.452 559 58 5	8 1.649 757 72 3.107 983 02 6	8 1.548 992 99 4.251 004 98 6	9 1.583 000 00 4.682 841 62 7
i		1.723 504 91 1.520 327 00	1.710 936 35 2.452 559 58	1.649 757 72 3.107 983 02	1.548 992 99 4.251 004 98	1.583 000 00 4.682 841 62

^aLine 1: number of basis function N_f .^bLine 2: spacing between exponential parameters β .^cLine 3: center basis function $\alpha\beta^{(N_f+1)/2}$ in atomic units.

carried out, the f -limit basis results are much better than the results obtained with the Dunning basis set, and as l_{\max} increases, the amount of improvement increases. At the largest basis for which we have results for all seven test species, the extrapolated f -limit basis results give a rms error that is a factor of 14 times smaller than those obtained from the extrapolated cc-pVXZ results. This clearly shows that the substitution $X = l_{\max}$ in Eq. (1), although reasonable, is not perfect.

We can also perform the extrapolation a little differently to obtain our best results. This is obtained by a least squares fit to the results of the three largest basis sets to Eq. (1) with $n=3$ for the singlet pair energies and (T) energies, and $n=5$ for the triplet pair energies. The fact that we are fitting three energies with two parameters stabilizes the results. This is important for the individual pair energies have two sources of error. The first is that in spite of all our efforts, the f -limit basis sets are not uniformly converged for each l . This will introduce errors of several μE_h . The second is we neglect transformed integrals smaller than a cutoff to decrease the

disk storage in the f -limit calculations. We expect this to cause errors of a fraction of a μE_h . In Table VI we give our results. The agreement with Klopper's results is very satisfactory. An independent check is the rms error of the least squares fit. The difference between our results and Klopper's results correlates well with this difference. The fits to the (T) energies are always very good, and the fits to the triplet pairs are very good, with the exception of F_2 , where results up to i basis functions are really required to obtain accurate results.

III. HOW EXTRAPOLATION CAN BE MADE TO WORK

We now wish to introduce more general formulas that include the X^{-n} dependence due to the Coulomb hole, but also recognize that our basis sets are not complete. We would also like to do this with minimal assumptions.

We will begin by assuming that we can write

$$E^c(X) = E^c(\infty) + \sum_n C_n^c f_n^c(X) \quad (5)$$

TABLE III. HF energies from f -limit basis sets (in E_h). The Ne energy is $-128.547\,092E_h$.

l_{\max}	N ₂	CH ₂	H ₂ O	CO	HF	F ₂
2	-108.989 985	-38.895 637	-76.065 841	-112.787 750	-100.069 531	-198.770 211
3	-108.992 871	-38.896 011	-76.067 402	-112.790 621	-100.070 810	-198.773 339
4	-108.993 061	-38.896 031	-76.067 455	-112.790 794	-100.070 847	-198.773 478
5	-108.993 072	-38.896 032	-76.067 457	-112.790 804	-100.070 849	-198.773 492
6	-108.993 074	-38.896 032		-112.790 805	-100.070 849	

TABLE IV. Singlet and triplet pair energies (in mE_h) from f -limit basis sets.

l_{\max}	Ne	N ₂	CH ₂	H ₂ O	CO	HF	F ₂
Singlet pair energies							
2	-163.685	-246.279	-122.500	-175.523	-235.073	-173.529	-340.648
3	-191.213	-269.049	-139.370	-196.611	-258.935	-198.169	-386.283
4	-202.637	-276.440	-141.677	-203.019	-267.054	-207.112	-402.604
5	-206.532	-279.101	-142.412	-205.086	-269.911	-210.095	-408.260
6	-208.248	-280.238	-142.714		-271.149	-211.385	
Triplet pair energies							
2	-95.347	-116.857	-34.351	-83.895	-114.082	-92.445	-170.525
3	-102.937	-123.867	-32.045	-89.524	-121.083	-99.159	-183.558
4	-104.333	-125.080	-32.290	-90.395	-122.308	-100.316	-185.973
5	-104.676	-125.369	-32.341	-90.587	-122.597	-100.588	-186.556
6	-104.791	-125.459	-32.356		-122.688	-100.674	

for some functions f_n^c . This form encompasses Eqs. (1) and (3) but not Eqs. (2) and (4), unless the nonlinear parameters a and b are somehow fixed. Nonetheless this is a very flexible form. The problem now is to determine the optimum functions. This is not at all easy, but we note that we do not really need to know the functions, but rather just how to determine $E^c(\infty)$. Thus we will circumvent the difficult task of finding the functions by noting that if the functions f_n are linearly independent, then Eq. (5) implies

$$E^c(\infty) = \sum_X E^c(X) F_X^c \quad (6)$$

subject to the constraint

$$1 = \sum_X F_X^c, \quad (7)$$

which arises from the first term on the right-hand side of Eq. (5). The constants F_X^c are determined solely by the f_n^c and not the values of $E^c(X)$, and so can be tabulated. Thus the task of determining the functions is changed to the task of determining the constants F_X^c , a much simpler task.

It should be noted that Varandas¹⁶ has also considered improved extrapolation formulas by including additional inverse powers in Eq. (1) with the additional coefficients being determined as functionals of the leading coefficient. This leads to violations of the constraint of Eq. (7). We believe that it is important to retain this constraint.

In order to proceed, we have to make an assumption about the functions f_n^c . Two choices come to mind: The first is to assume that they are universal, i.e., they neither depend on molecular species nor on molecular geometry. The second choice is they depend on the molecular species, but not on the molecular geometry. In either case, if we have a database

TABLE V. (T) energies from f -limit basis sets (in mE_h).

l_{\max}	Ne	N ₂	CH ₂	H ₂ O	CO	HF	F ₂
2	-4.918	-18.449	-4.726	-8.127	-16.986	-7.059	-18.903
3	-6.008	-20.480	-5.432	-9.407	-18.822	-8.323	-21.697
4	-6.296	-20.966	-5.578	-9.686	-19.270	-8.627	-22.423
5	-6.401	-21.132	-5.619	-9.772	-19.425	-8.729	-22.673
6	-6.442	-21.199	-5.635		-19.487	-8.768	

of accurate estimates of $E^c(\infty)$, we can invert Eq. (6) via a least squares fit to determine the F_X^c . This is the crux of our method.

In this paper we will assume that the functions are universal, and will use Klopper's CCSD-R12 results¹⁰ for the seven species to determine the F_X^c for the CCSD energies and our (T) results for the (T) extrapolation. It should be noted that the extrapolation coefficients will depend on the family of basis sets used, e.g., cc-pVXZ vs aug-cc-pVXZ. In Table VII we give the coefficients we have determined. For comparison, we also give the coefficients obtained using the power law of Eq. (1).

The simplest case is when we consider the results of two basis sets, in which case we have a single parameter to optimize in the expression

$$E^c(\infty) = [E^c(X_2) - E^c(X_1)] F_{x_2}^c + E^c(X_1). \quad (8)$$

In Fig. 2 we compare the various CCSD extrapolations. We find the very interesting results that the best "conventional" extrapolation is obtained using aug-cc-pVXZ basis sets and X^{-3} for singlet pairs and X^{-5} for triplet pairs, while

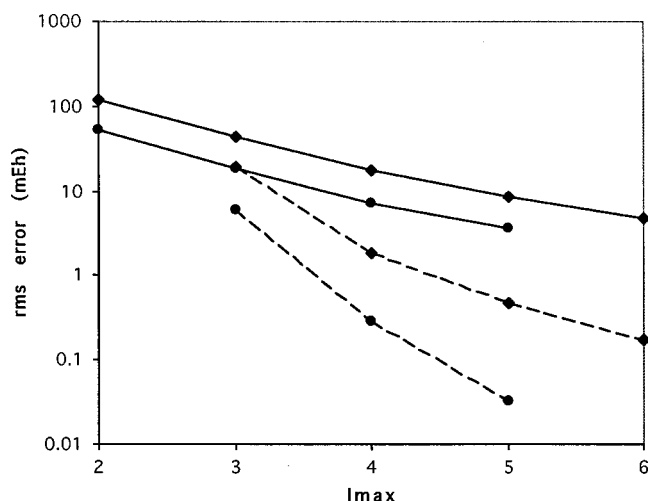


FIG. 1. Root-mean-square errors from cc-pVXZ and f -limit basis sets. Solid lines are unextrapolated, dashed lines are extrapolated using Eq. (1) with l_{\max} and $l_{\max}-1$ for singlet and triplet pairs separately, diamonds are the cc-pVXZ basis results, and circles are the f -limit basis results.

TABLE VI. Best estimates of correlation energies (in mE_h).

	Ne	N ₂	CH ₂	H ₂ O	CO	HF	F ₂
Singlet pair energies							
pw ^a	-210.61	-281.85	-143.16		-272.88	-213.20	
Klopper ^b	-210.61	-281.88	-143.16		-272.92	-213.14	
Triplet pair energies							
pw	-104.85	-125.51	-32.37	-90.67 ^c	-122.74	-100.73	-186.78 ^c
Klopper ^b	-104.87	-125.53	-32.37	-90.69	-122.75	-100.74	-186.87
(T) energies							
pw	-6.505	-21.300	-5.660	-9.878 ^c	-19.580	-8.830	-22.945 ^c

^aPresent work.^bReference 10.^cUsing only up to h functions.

in contrast, all of the new two-point extrapolation parameters give essentially identical results for either basis set, and furthermore, these results are significantly more accurate than the power law extrapolations. That aug-cc-pVXZ extrapolations are more reliable was also observed by Valeev *et al.*¹⁷ When using the power law extrapolation, results roughly equivalent to $X=Q$ are obtained from $X=D$ and T . In contrast, when using the new formula, $X=D$ and T yields results roughly equivalent to $X=7$. As the largest X used in the extrapolation increases, the power law and new formulas give more similar results, but the new formula always gives significantly more accurate results. We only show one curve for the new results as the curves obtained using the cc-pVXZ and aug-cc-pVXZ basis sets are indistinguishable. This is not the case for the power law extrapolations.

It is interesting to compare the present results to the extrapolation of Truhlar.¹⁸ Truhlar optimized the exponent in Eq. (1) to give the best agreement with accurate results using $X=D$ and T . Since he intended that the exponent be used only for $X=D$ and T , functionally his method is equivalent to optimizing the coefficient in Eq. (8), but our approach has some advantages: First of all, the process of determining the optimum coefficient is much easier numerically. Second, we have a simple, well defined, procedure for including more terms in Eq. (6). The final curve in Fig. 2 illustrates this by showing the effect of extrapolating using three values of X . The rms error is reduced by about a factor of 2 for a given largest value of X compared to the two-point formulas.

In Fig. 3 we compare various methods of extrapolating the (T) energies for both the cc-pVXZ and aug-cc-pVXZ basis sets. Here we find that extrapolating the (T) energies with X^{-3} leads to erratic results, while the fitted coefficients give very nice results, and furthermore, the difference between the cc-pVXZ and aug-cc-pVXZ results is not that great. The results obtained with the new coefficients are always at least an order of magnitude more accurate than not extrapolating.

In Fig. 4 we show a comparison of extrapolating the SCF energy via Eq. (4) or by Eq. (6). We find the exponential extrapolation to be very erratic—sometimes producing rms errors greater than the inextrapolated results; however, in contrast, the new extrapolation works very well, always considerably decreasing the rms error.

IV. DISCUSSION

In this paper we have introduced extrapolation formulas for CCSD(T) energies that are much more accurate than previous ones. We find that it is not necessary to decompose the CCSD energy into singlet pair and triplet pair contributions to obtain accurate results. This is significant for two reasons: First of all, not all CCSD codes form this decomposition. Second, it is not possible to carry out this decomposition for open-shell CCSD calculations, because the wave function is not a spin eigenfunction in practical implementations of open-shell CCSD theory. We also find that equivalent results are obtained with both the cc-pVXZ and aug-cc-pVXZ basis sets.

TABLE VII. Coefficients for extrapolation.

X_1 and X_2	Power	cc-pVXZ	aug-cc-pVXZ
SCF			
2 and 3		1.332 527 6	1.347 630 2
3 and 4		1.307 126 9	1.294 053 1
4 and 5		1.144 266 6	1.109 913 7
5 and 6		1.204 123 2	1.119 855 0
Singlet pairs CCSD			
2 and 3	1.421 053	1.707 912 0	1.694 220 2
3 and 4	1.729 730	1.767 411 9	1.759 252 4
4 and 5	2.049 180	1.987 349 7	2.005 973 6
5 and 6	2.373 626	2.316 158 3	2.333 172 0
Triplet pairs CCSD			
2 and 3	1.1516 59	1.356 600 5	1.331 348 8
3 and 4	1.3111 40	1.464 094 4	1.454 067 5
4 and 5	1.4873 87	1.518 271 4	1.529 966 8
5 and 6	1.671 899	1.742 258 9	1.755 288 6
Total CCSD			
2 and 3		1.595 712 1	1.587 761 6
3 and 4		1.699 881 4	1.700 111 5
4 and 5		1.900 400 2	1.930 317 4
5 and 6		2.237 550 1	2.265 620 6
(T)			
2 and 3		1.503 285 2	1.398 597 3
3 and 4		1.695 134 7	1.730 158 4
4 and 5		1.741 321 2	1.810 472 6
5 and 6		2.101 801 0	2.247 961 7

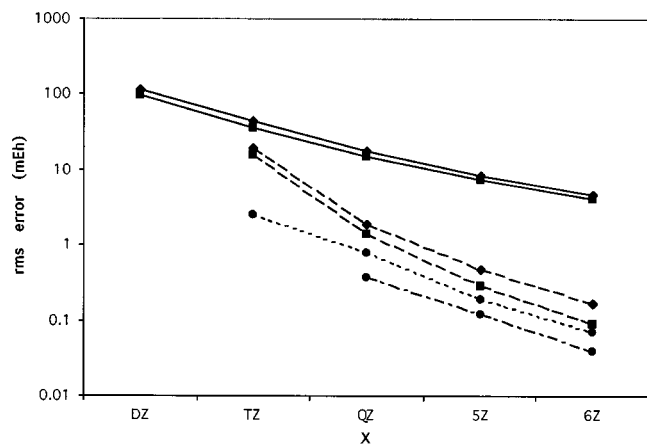


FIG. 2. Rms errors of CCSD energies. Diamonds denote cc-pVXZ basis sets, squares denote aug-cc-pVXZ basis sets, and circles both. Solid lines mean no extrapolation; dashed line, extrapolation using power laws using X and $X-1$ for singlet and triplet pair energies separately; dotted line, two point extrapolation using the new formula with X and $X-1$; and dash-dotted line, three point extrapolation using the new formula with X , $X-1$, and $X-2$.

The extrapolation of all components of the energy is linear; thus the extrapolations can also be easily used with analytic derivatives of the energy.

An important question is how extendible the extrapolation parameters are to other systems. We have attempted to answer this question by carrying out fits that only include six of the seven test systems, and using the omitted system as a test of the extendibility of the coefficients. We carry out the fits omitting each test system in turn and computing the error of the omitted system. The rms error of the omitted systems is then compared to the rms error obtained from fitting all systems. The ratio of these errors range from 1.1 to 1.5, depending on the energy being extrapolated, the basis type, and the X involved in the fit. For the singlet pair energies, which have the slowest convergence, the ratio ranges from 1.1 to 1.2. Thus we conclude that extrapolation parameters we produce have good extendibility.

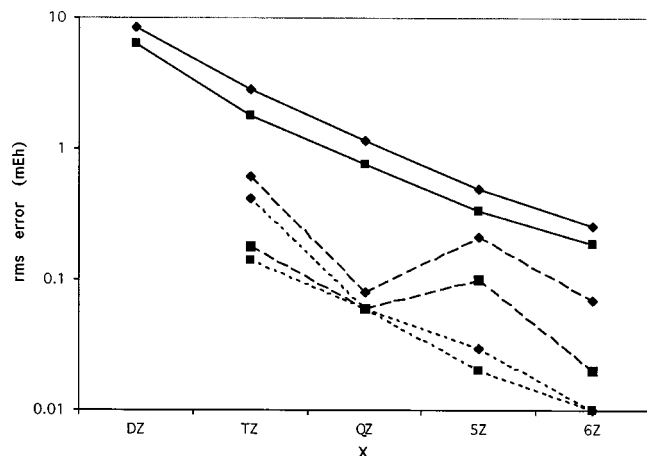


FIG. 3. Rms errors from extrapolation of (T) energies. Diamonds are cc-pVXZ basis results, squares aug-cc-pVXZ results, solid lines, unextrapolated, dashed lines, extrapolated using power law with X and $X-1$, dotted, extrapolated using new formula with X and $X-1$.

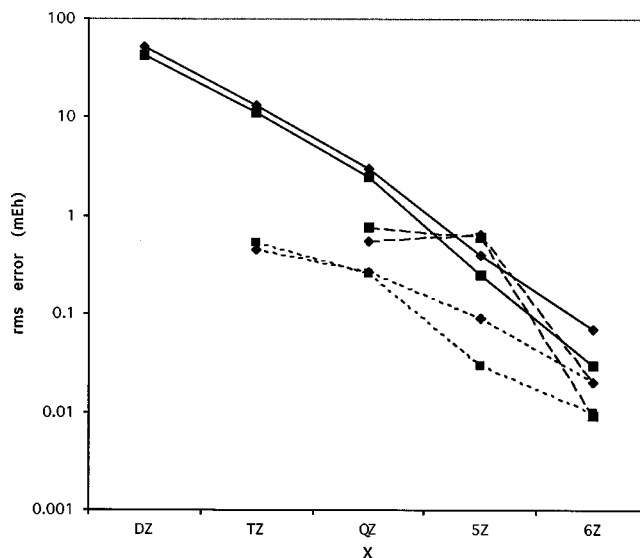


FIG. 4. Rms errors for extrapolation of SCF energies. Diamonds denote cc-pVXZ basis results and squares aug-cc-pVXZ basis results. Solid lines are unextrapolated, dashed lines are extrapolated using the exponential form Eq. (4) with X , $X-1$, and $X-2$, and dotted are extrapolating using the new formula using X and $X-1$.

It is worth considering the special case H_2 . Since $X = l_{\max} - 1$ for H rather than $X = l_{\max}$ for the first row elements, it seems likely that a different set of extrapolation parameters might be required. We have carried out test calculations at $r = 1.4a_0$, and we find that the new extrapolation formulas work just as well for H_2 as for the seven test species, in spite of the fact that the TP contribution to the correlation energy is zero.

Although we do not consider weakly bound systems in this work, there are other works in the literature that suggest that the extrapolation schemes suggested here will work well for weakly bound systems. Park and Lee¹⁹ found that for dimers of He, HF, and H_2O , accurate extrapolation to the basis set limit could be obtained by using extrapolation rules optimized for the monomers. However, compared to the present work, they used different coefficients for each system.

We also do not consider the extrapolation of the core-valence correlation energy in the present work. Klopper¹⁰ gives values for the CCSD-R12 core-valence correlation for the seven test species, and it would be straightforward to apply the present method to deduce extrapolation coefficients. However, since the energies Klopper gives includes the core correlation energy, it is not clear how to preferentially extrapolate the core-valence energy, which is the energy of primary interest.

¹G. D. Purvis and R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982).

²J. Almlöf and P. R. Taylor, J. Chem. Phys. **86**, 4070 (1987).

³T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

⁴C. Schwartz, Phys. Rev. **126**, 1015 (1962); C. Schwartz, Methods Comput. Phys. **2**, 241 (1963).

⁵W. Kutzelnigg and J. D. Morgan, J. Chem. Phys. **96**, 4484 (1992).

⁶J. M. L. Martin, Chem. Phys. Lett. **259**, 669 (1996).

⁷K. A. Peterson, D. E. Woon, and T. H. Dunning, J. Chem. Phys. **100**, 7410 (1994).

⁸D. Feller, J. Chem. Phys. **98**, 7059 (1993).

- ⁹A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, *Chem. Phys. Lett.* **302**, 437 (1999).
- ¹⁰W. Klopper, *Mol. Phys.* **99**, 481 (2001).
- ¹¹D. E. Woon and T. H. Dunning, *J. Chem. Phys.* **103**, 4572 (1995).
- ¹²D. E. Woon and T. H. Dunning, *J. Chem. Phys.* **100**, 2975 (1994).
- ¹³R. D. Bardo and K. Ruedenberg, *J. Chem. Phys.* **60**, 918 (1974).
- ¹⁴W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes* (Cambridge University Press, Cambridge, 1992).
- ¹⁵H.-J. Werner and P. J. Knowles, MOLPRO, a package of *ab initio* programs with contributions from J. Almlöf, R. D. Amos, A. Berning *et al.*, <http://www.molpro.net>.
- ¹⁶A. J. C. Varandas, *J. Chem. Phys.* **113**, 8880 (2000).
- ¹⁷E. F. Valeev, W. D. Allen, R. Hernandez, C. D. Sherrill, and H. F. Schaefer, *J. Chem. Phys.* **118**, 8594 (2003).
- ¹⁸D. G. Truhlar, *Chem. Phys. Lett.* **294**, 45 (1998).
- ¹⁹S. Y. Park and J. S. Lee, *J. Chem. Phys.* **116**, 5389 (2002).