Theoret. Chim. Acta (Berl.) 52, 231–251 (1979)

THEORETICA CHIMICA ACTA

Systematic Approach to Extended Even-Tempered Orbital Bases for Atomic and Molecular Calculations*

David F. Feller** and Klaus Ruedenberg

Ames Laboratory, USDOE, and Department of Chemistry, Iowa State University, Ames, Iowa 50011, USA

Explicit formulas are established for simply generating arbitrarily large basis sets of optimal even-tempered Gaussian primitives which systematically approach complete bases for the entire function space. These bases, moreover, reproduce the corresponding optimal atomic SCF wavefunctions extremely closely and permit an extrapolation of the SCF energies to the Hartree–Fock limit. On the basis of the detailed quantitative information available from these calculations a simple general procedure is formulated for generating optimal even-tempered basis sets for molecular calculations.

Key words: Extended even-tempered orbital bases

1. Introduction

If optimal performance is to be achieved with Gaussian basis sets, which are necessarily restricted in size due to computational limitations, then a good understanding of the relationship between such sets and complete bases is needed in order to gain an accurate assessment of convergence rates for the expectation values derived from variational calculations. About 15 years ago Schwartz [1] wrote, "The first essential in talking of convergence rates is to have an orderly plan of procedure. That is, one must choose a set of basis functions to be used and then gradually add more and more of these terms to the variational calculation in some systematic manner."

---

* Dedicated to Professor Dr. Hermann Hartmann on the occasion of his 65th birthday.

** Present address: Chemistry Department, University of Washington, Seattle, Washington.

0040–5744/79/0052/0231/$04.20

The even-tempered bases introduced by Ruedenberg *et al.* [2] are ideally suited for this purpose. Even-tempered (ET) Gaussian primitives are defined in terms of two optimizable parameters per symmetry by

$$\chi(klm) = N(\zeta_k) \exp(-\zeta_{kl} r^2) r^l S_l^m(\theta, \phi), \quad k = 1, 2, \ldots, \tag{1a}$$

$$\zeta_{kl} = \alpha_l \beta_l^k. \tag{1b}$$

While the restriction of the orbital exponents to a geometric sequence entails a small loss in variational freedom when compared to an independently optimized set, this loss must be seen in perspective. For example, the first row elements might require an additional *s*-type primitive or two to obtain groundstate energies within a few millihartrees of Huzinaga's [3] completely optimized exponent set for an (8*s*, 4*p*) basis, but the difference between the best energy obtainable with this set and the infinite basis set limit is several times larger. Table 1 shows the actual SCF energies for the (8*s*, 4*p*) ET and independently optimized atomic basis sets on carbon. All energies in this table and in the remainder of the paper are in atomic units (1 a.u. = 1 hartree = 627.5 Kcal/mole).

For this particular basis the ET choice of exponents has resulted in a 3/1 reduction in the number of nonlinear parameters, the independent exponents, which have to be varied in the optimization procedure. In larger basis sets the savings are even greater.

It is the object of this investigation to show that even-tempered Gaussian basis sets indeed furnish effective vehicles for systematically approaching complete basis sets. We begin by demonstrating quantitative relationships between even-tempered expansions and certain integral transform representations of exact Hartree–Fock atomic orbitals. Next, we establish a simple procedure for constructing appropriate sequences of even-tempered orbital sets of increasing size for several illustrative atoms and use them to systematically approach atomic Hartree–Fock limits. On the basis of this information we finally develop a simple general method for constructing atomic orbital basis sets, called ETGSCFD-type, that are effective for molecular calculations.

Table 1. SCF energies for the carbon atom with ET and independently optimized Gaussian bases

Basis	Energy	ΔE_∞ [a]	ΔE_H [b]
Huzinaga (8*s*, 4*p*)	−37.6798	0.0088	0.0
ET (8*s*, 4*p*)	−37.6681	0.0205	0.0117
ET (9*s*, 4*p*)	−37.6768	0.0118	0.0030
ET (23*s*, 11*p*)	−37.6886	0.0000	−0.0088

[a] ΔE_∞ is the difference between the energy for this basis and the HF limit energy.

[b] ΔE_H is the difference between the energy of this basis and the Huzinaga (8*s*, 4*p*) basis result.

2. Even-Tempered Gaussian and Exponential Expansions and Gaussian Integral Transforms for Atomic Orbitals

In order to elucidate the relationship between exact HF-AO's and their approximations in terms of even-tempered Gaussian AO's, we note that any atomic orbital of symmetry (l, m) can be expressed through an integral transform over Gaussian radial functions in the following manner

$$\phi_{lm}(r, \theta, \phi) = S_l^m(\theta, \phi) r^l \int_0^\infty d\zeta \, e^{-\zeta r^2} f_l(\zeta), \tag{2}$$

where $S_l^m(\theta, \phi)$ is a normalized spherical harmonic. In the present context it is convenient to write this representation in the form

$$\phi_{lm}(r, \theta, \phi) = \int_{-\infty}^\infty d(\ln \zeta) g_{lm}(\zeta; r, \theta, \phi) a_l(\zeta), \tag{3}$$

where

$$g_{lm}(\zeta; r, \theta, \phi) = N_l \, e^{-\zeta r^2} r^l S_l^m(\theta, \phi), \tag{4a}$$

with

$$N_l = \{\zeta^{2l+3} 2^{4l+7} / \pi[(2l+1)!!]^2\}^{1/4}, \tag{4b}$$

is a normalized Gaussian-type primitive AO. If we consider for ϕ_{lm} exponential-type normalized AO's of the form

$$\chi(\xi; r, \theta, \phi) = M_l \, e^{-\xi r} r^l S_l^m(\theta, \phi), \tag{5a}$$

with

$$M_l = \{(2\xi)^{2l+3} / (2l+2)!\}^{1/2}, \tag{5b}$$

then the transform function a_l in Eq. (3) becomes

$$a_l^{\text{ex}}(\xi, \zeta) = [2^{l+1}/(l+1)! \sqrt{\pi/2}]^{1/2} (\xi^2/4\zeta)^{(2l+5)/4} \, e^{-(\xi^2/4\zeta)}. \tag{6}$$

This is a generalization of a formula given by Kikuchi [4] for simple exponentials. Bishop and Somorjai [5] as well as Taylor [6] have also examined transforms of radial functions.

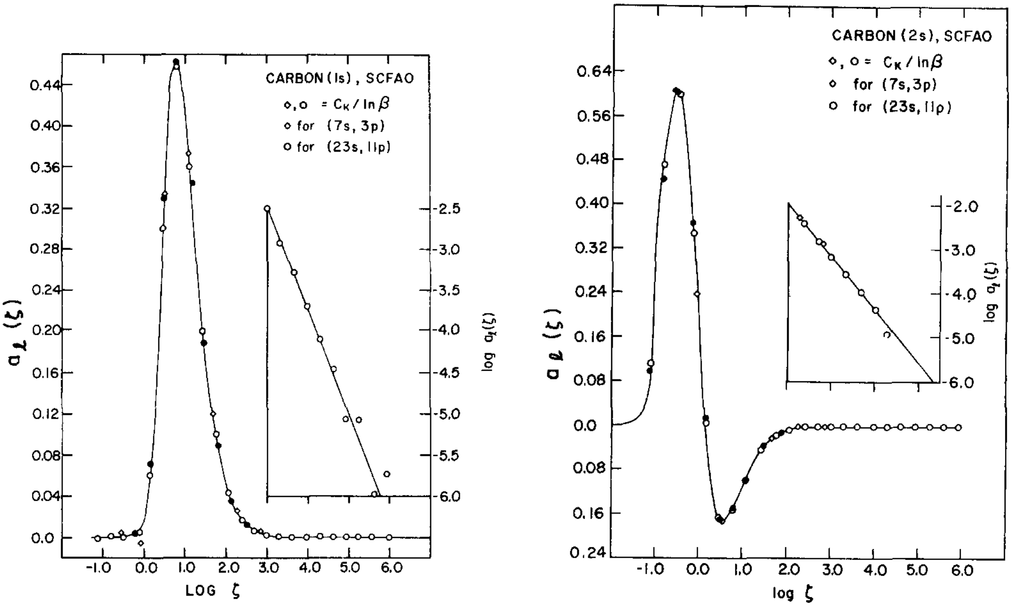
It has been shown by Raffenetti [7] that any HF SCF AO can be efficiently expanded in terms of even-tempered exponential-type AO's of the form (5a),

$$\phi_{lm} = \sum_v b_v \chi_{lm}(\xi_{lv}), \quad \xi_{lv} = \hat{\alpha}_l \hat{\beta}_l^v. \tag{7}$$

Combining such an expansion with the integral transforms for χ_{lm}, we find for ϕ_{lm} the transform function

$$a_{lm}(\zeta) = \sum_v b_v a_l^{\text{ex}}(\xi_{lv}, \zeta), \tag{8}$$

where a_l^{ex} is given by Eq. (6). Examples of such HF AO transforms are shown in Figs. 1a–1c. Fig. 1a corresponds to the $(1s)$ orbital, Fig. 1b to the $(2s)$ orbital and Fig. 1c to the $(2p)$ orbital of the HF wavefunction of the carbon groundstate. The



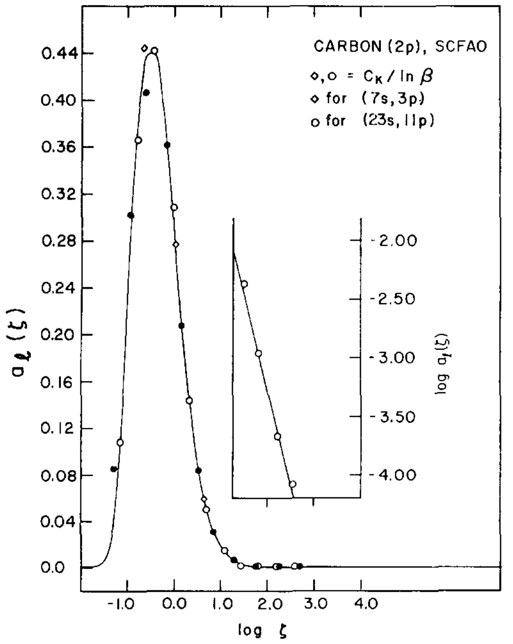


Fig. 1(a) Integral transform and ETG expansion coefficients of the 1*s* orbital in the Carbon groundstate. Solid dots = flourine expansion coefficients (see text). **(b)** Integral transform and ETG expansion coefficients of the 2*s* orbital in the Carbon groundstate. Solid dots = fluorine expansion coefficients (see text). **(c)** Integral transform and ETG expansion coefficients of the 2*p* orbital of the Carbon groundstate. Solid dots = fluorine expansion coefficients (see text) values for the b_ν and $\xi_{l\nu}$ are taken from Raffenetti's (6s, 4p) even-tempered exponential expansion. Since this is an extremely accurate wavefunction (triple zeta in s and double zeta in p) the curves in these three figures can be considered as very close to the Gaussian transforms of the exact carbon HF SCF orbitals.

Approximation of the integral transform (3) by means of a numerical integration implies replacement of the integral by a sum over a number of grid points ζ_k ($k = 1, 2, 3 \ldots$). Since it is apparent that the intervals ($\zeta_{k+1} - \zeta_k$) should increase as ζ_k becomes larger and larger, one reasonable choice of gridpoints is given by the even-tempered exponents introduced in Eq. (1b), namely $\zeta_k = \alpha_l\beta_l^k$. This choice leads to a set of *equidistant* gridpoints when ln (ζ) is chosen as the integration variable, as has been implied in Eq. (3). Since the distance between neighboring gridpoints (ln ζ_k) is ln (β_l), the even-tempered grid approximation to the integral transform (3) thus becomes

$$\phi_{lm} \approx \sum_k g_{lm}(\zeta_k) a_l(\zeta_k) \ln \beta_l. \tag{9}$$

This type of approximation for HF AO's in terms of g_{lm} can be compared to those expansions that result from direct HF SCF calculations based on expansions of the SCF AO's in terms of even-tempered Gaussian primitives AO's, *viz.*

$$\tilde{\phi}_{lm} = \sum_k g_{lm}(\zeta_k) c_k^l, \quad \zeta_k = \alpha_l \beta_l^k, \tag{10}$$

where c_k^l as well as the α_l and β_l^k are variationally determined. In view of Eqs. (8), (9), and (10) one would expect relations like

$$c_k^l \approx a(\alpha_l\beta_l^k) \ln \beta_l \approx \ln \beta_l \sum_\nu b_\nu^l a_l^{\mathrm{ex}}(\xi_\nu, \zeta_k = \alpha_l\beta_l^k), \tag{11}$$

in which b_ν^l, ξ_ν come from HF calculations in terms of exponentials, whereas c_k^l, α_l, β_l come from HF calculations in terms of Gaussians. The a_l^{ex} are given by Eq. (6).

As a first example we consider the hydrogen (1s) function $(\xi^3/\pi)^{1/2} e^{-\xi r}$. In this case Eq. (11) simplifies to

$$c_k \approx (8/\pi)^{1/4} \ln \beta (4\alpha\beta^k)^{-5/4} e^{-1/4\alpha\beta k}. \tag{12}$$

If the c_k, α, β are determined from a six-term even-tempered Gaussian expansion (corresponding to an error in the total energy of 0.2 millihartrees), then the agreement of the left and right hand sides of (12) is better than two significant figures. For an eight-term expansion (corresponding to an energy error of 0.01 millihartrees) the agreement is better than four significant figures.

Next we consider the carbon groundstate HF AO's whose integral transforms were shown in Figs. 1a–1c. Specifically we choose two sets of parameters with one set resulting from an SCF calculation using a "small" basis of even-tempered Gaussian primitives and the second set resulting from an SCF calculation with a "large" basis of even-tempered primitives, the former being a (7s, 3p) basis, the latter a (23s, 11p) basis. In order to test the degree of validity of Eq. (11) we simply plot for the three atomic orbitals the values of the quantities (c_k^l/ln β_l) for the appropriate abscissa values of $\zeta_k = \alpha_l \beta_l^k$ on the curve for a $a_l(\zeta_k)$. The (7*s*, 3*p*) values are entered as diamonds, the (23*s*, 11*p*) values as circles. It is apparent that not only is the agreement perfect for the large basis, but it is also very good for the small basis. This agreement between the direct variational coefficients c_k^l and the exact transform functions shows that the variational representation in terms of even-tempered Gaussian primitives approaches the exact SCF solution in a systematic manner. The integral transform acts as if it were a "slidewire" with the coefficients for finite expansions behaving like beads on the wire. The gridpoints of a particular expansion merely determine the positions of the beads on the wire and these positions can be altered without departing from the wire. We have noticed similar behavior for expansions of molecular orbitals.

Another interesting aspect of the discussed results is that they establish a clear relation between the expansion of an atomic orbital in terms of Gaussian primitives and its expansion in terms of exponential-type primitives. Equation (11) shows how to obtain the coefficients of the even-tempered Gaussian expansion when the Gaussian and exponential exponents and the exponential coefficients are known. Thus, using this equation it was possible to predict the coefficients of a 9-term Gaussian expansion of the 1*s* and 2*s* orbitals of carbon with sufficient accuracy that an energy only 0.009 E_h above the SCF coefficients was produced without actually performing the SCF calculation with Gaussians. However, it is also possible to invert the process. Since Gaussian expansions are always substantially longer than exponential-type expansions of equal quality, it is apparent that the number of coefficients c_k^l for which Eq. (11) applies is larger than the number of terms in the summation over ν. If it is more than twice as large, then there are sufficient equations available to determine the values of the parameters for the exponential expansion when the Gaussian parameters c_k^l, α_l, β_l, are known. This can be accomplished by a (partly linear, partly nonlinear) least-squares calculation based on minimizing the quantity

$$\sum_k [c_k^l - f^l(b_1^l, b_2^l, \ldots; \xi_1, \xi_2, \ldots)]^2$$

with respect to the parameters b_ν^l and ξ_ν. Here the f_k^l are the functions defined by the right hand side of Eq. (11) and the values of c_k^l, α_l, β_l are supplied by the Gaussian expansion. It is thus possible to deduce the complete even-tempered exponential-type expansion from sufficiently large even-tempered Gaussian expansions. This procedure works quite well, as the results shown in Table 2 for the carbon 1*s* and 2*s* orbitals attest.

The shape of the Gaussian transform is not only highly independent of the basis set size, it is also nearly independent of the atomic number. Of course, the shrinkage of the atomic orbitals with increasing nuclear charge results in a shift of the transform function to higher log ζ values for larger Z. However, by suitable shifting of the transforms of two different elements, they can be superimposed almost perfectly. This is also shown in Figs. 1a–1c, where the values of $(c_k^l/\ln \beta_l)$ from a (22*s*, 11*p*) SCF calculation of fluorine, shifted into maximal alignment with the carbon transform, are indicated as solid black dots. The magnitude of the shift

Table 2. Carbon SCF orbital parameters for a (6*s*, 4*p*) even-tempered exponential basis. Comparison between the parameters obtained from a direct SCF calculation in the ETE basis with those obtained by fitting to orbitals from an SCF calculation in terms of a (19*s*, 9*p*) ETG basis[a]

Parameters obtained by fitting to orbitals from SCF calculation in terms of ETG basis.

$a = 0.702, b = 1.666$

	$C(1)$	$C(2)$	$C(3)$	$C(4)$	$C(5)$	$C(6)$
1*s*	−0.003	0.010	−0.002	0.922	0.080	0.002
2*s*	−1.250	−0.060	0.524	0.118	−0.004	0.007

Parameters obtained from SCF calculation in terms of ETE basis.

$\hat{a} = 0.705, \hat{b} = 1.667$

	$C(1)$	$C(2)$	$C(3)$	$C(4)$	$C(5)$	$C(6)$
1*s*	−0.000	0.002	0.016	0.913	0.077	0.001
2*s*	−1.252	−0.062	0.545	0.092	−0.011	0.001

[a] Both basis sets gave an energy of $-37.68859\ E_h$. The parameters a and b were determined by nonlinear fitting, while $\hat{a}$ and $\hat{b}$ are the SCF optimized even-tempered exponential values of Raffenetti [7].

is approximately 0.4 for both s and p symmetries. To avoid confusion, only the fluorine data points near the maximum and minimum are plotted for the 1*s* and 2*s*; the agreement is no less good for high and low ζ values. Most surprising is the excellent agreement for the 2*p* AO in spite of the additional three electrons using that radial orbital factor in fluorine.

3. Regularities in the Optimal Atomic ETG Exponential Parameters

How do the optimal atomic (α, β) depend on the expansion length used in an atomic SCF calculation? In order to answer this question, several basis sets ranging in size up to 16 *s*-type and 7 *p*-type Gaussian primitives for the first row elements and up to 9 *p*-type primitives for sulfur were optimized by varying (α, β) until the lowest energy for the appropriate groundstate was obtained. For the s symmetry optimizations, four *p*-primitives were used for the 2*p* AO in C and O, while six *p*-primitives were used in S and Se. For the p symmetry seven *s*-primitives were used for the 1*s* and 2*s* AO's in C and O, while eight *s*-primitives were used in S and Se. It was established that the optimization of one symmetry is highly independent of the number of primitives in the other symmetries so long as the other symmetry is not overly truncated. Table 3 gives the optimal values found.

In Figs. 2a–2c the ln (ln (β)) for some of the optimal parameters listed in Table 3 is seen to be linearly related to the ln $(N - 1)$ where N is the number of primitives of that symmetry in the basis set. This linear dependence is suggested by the behavior of the largest exponent, i.e. $\lim (\alpha\beta^N) = \infty$ and the smallest exponent, i.e. $\lim \alpha\beta = 0$, whence $\lim \beta^{N-1} = \infty$ and hence $\lim [(N - 1) \ln \beta] = \infty$, whereas $\lim \beta = 1$ and hence $\lim (\ln \beta) = 0$. Thus $\ln \beta$ must tend to zero less strongly than

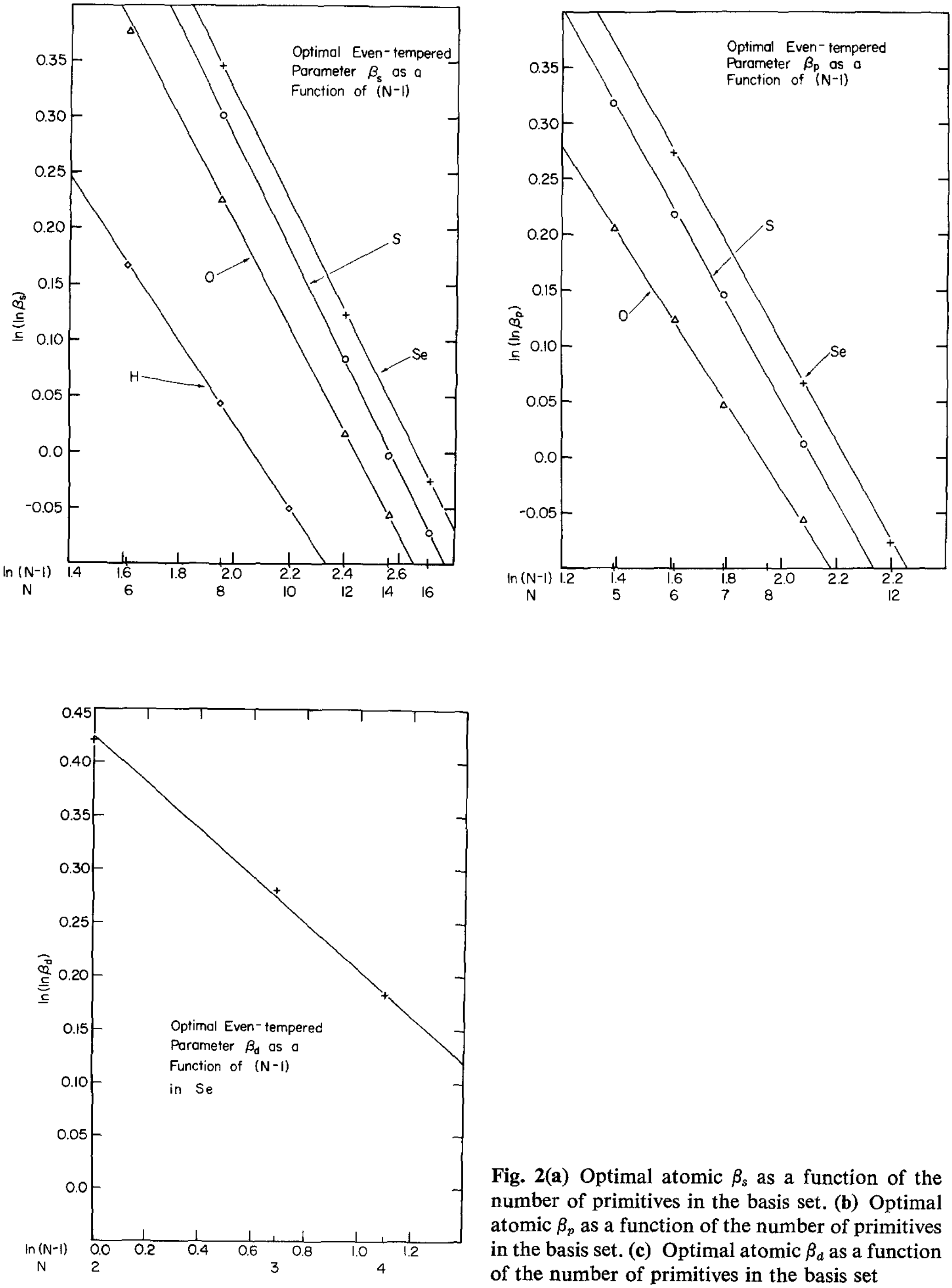


Fig. 2(a) Optimal atomic β_s as a function of the number of primitives in the basis set. **(b)** Optimal atomic β_p as a function of the number of primitives in the basis set. **(c)** Optimal atomic β_d as a function of the number of primitives in the basis set

Table 3. Optimal ET parameters for ground state carbon, oxygen, sulfur and selenium

	Carbon		Oxygen		Sulfur		Selenium	
Ns	Alpha	Beta	Alpha	Beta	Alpha	Beta	Alpha	Beta
5	0.05813	5.0784	–	–	–	–	–	–
6	0.06304	4.3406	0.11979	4.2931	–	–	–	–
7	0.05701	3.7973	0.10763	3.7710	–	–	–	–
8	0.05090	3.5089	0.09469	3.4995	0.07515	3.8708	0.10891	4.1055
9	0.05019	3.2937	0.09357	3.2790	0.07917	3.5992	–	–
10	0.04989	3.1112	0.09311	3.0895	0.07982	3.3199	–	–
11	0.04798	3.9032	0.08874	2.8978	0.06601	3.0920	–	–
12	0.04495	3.7647	0.07982	2.7649	0.06442	2.9667	0.08420	3.1013
13	–	–	–	–	0.06330	2.8338	–	–
14	0.04339	3.5790	0.08032	2.5709	0.06056	2.7079	–	–
15	–	–	–	–	0.05901	2.6143	–	–
16	–	–	–	–	0.05829	2.5385	0.06852	2.6510

	Carbon		Oxygen		Sulfur		Selenium	
Np	Alpha	Beta	Alpha	Beta	Alpha	Beta	Alpha	Beta
3	0.04550	4.4504	0.08159	4.5997	–	–	–	–
4	0.04168	3.7920	0.07205	3.8890	–	–	–	–
5	0.03806	3.3503	0.06523	3.4182	0.07202	3.9473	–	–
6	0.03523	3.0382	0.05859	3.1034	0.06446	3.4728	0.12810	3.7250
7	0.03229	2.8151	0.05477	2.8553	0.05263	3.1757	–	–
8	–	–	–	–	0.04859	2.9505	–	–
9	–	–	–	–	0.04488	2.7523	0.06941	2.9142
12	–	–	–	–	–	–	0.04422	2.5211

	Selenium	
Nd	Alpha	Beta
2	0.68011	4.5820
3	0.57640	3.7667
4	0.44612	3.3172

$(N - 1)$ tends to infinity. This would suggest a function of the form $\ln \beta \sim (N - 1)^k$ with $-1 < k < 0$, whence

$$\ln(\ln(\beta)) = k \ln(N - 1) + l. \tag{13}$$

Figure 3 shows the approximately linear dependence of $\ln(\alpha)$ on $\ln(\beta)$. For the larger bases this is seen to give a very close fit to an equation of the form

$$\ln(\alpha) = m \ln(\beta) + n. \tag{14}$$

The values for the constants appearing in Eqs. (13) and (14), which are obtained by least squares fitting the data for H, C, O, S and Se, are listed in Table 4. It would seem a difficult task to *a priori* predict the parameters of Eqs. (13) and (14) from simple analytical reasoning. The slopes of $\ln(\ln(\beta))$ as a function of $\ln(N - 1)$ can

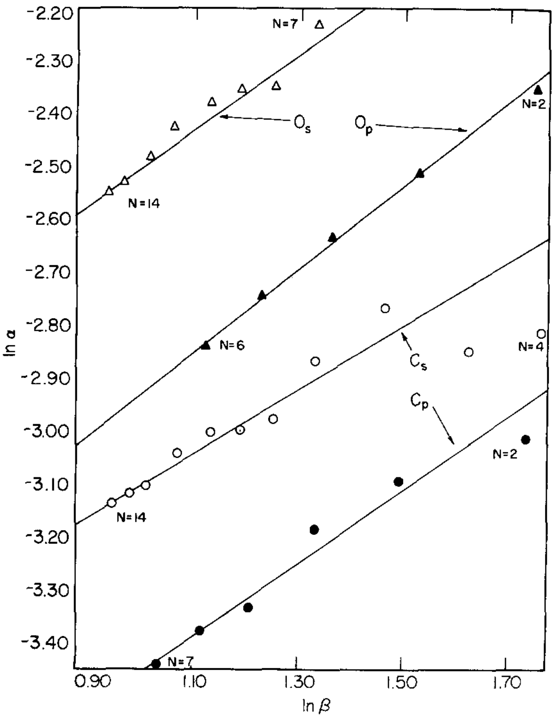


Fig. 3. Optimal ETG atomic parameters: ln α_l as a function of β_l

be found by an r-weighted least-squares fitting of accurate exponential basis set atomic calculations [8], but the intercept of this line and both the slope and intercept of the ln (α) vs. ln (β) line cannot be accurately determined in this manner.

Energy losses encountered when optimal (α, β) values are replaced with those values predicted from Eqs. (13) and (14) vary with basis sets and are listed in Table 5. By comparing the values in this table with the optimal ET exponential results of Raffenetti [7] for the same atoms, it can be seen that the ratio of the number of Gaussians to exponentials required to achieve the same total energy is approximately 3:1 for the first row, 2.4:1 for the second row, and 2:1 for the third row of

Table 4. Constants in the straight line approximations for the ET parameters

Atomic number	s-symmetry				p-symmetry			
	k	l	m	n	k	l	m	n
1	−0.369	0.763	0.467	−3.983				
6	−0.465	1.084	0.704	−3.810	−0.362	0.689	0.708	−4.163
8	−0.443	1.084	0.769	−3.269	−0.362	0.685	0.820	−3.760
16	−0.487	1.278	0.711	−3.509	−0.440	0.929	1.505	−4.640
34	−0.485	1.290	–	–	−0.443	0.988	–	–

Table 5. Optimal and "straight line" atomic energies

	Carbon			Oxygen	
Basis	Optimal	Str. line	Basis	Optimal	Str. line
(7*s*, 3*p*)	−37.630142	−37.628856	(5*s*, 3*p*)	−74.343343	−74.330671
(9*s*, 4*p*)	−37.676799	−37.676754	(7*s*, 4*p*)	−74.716530	−74.715517
(11*s*, 5*p*)	−37.685532	−37.685492	(9*s*, 5*p*)	−74.791194	−74.791182
(13*s*, 6*p*)	−37.687815	−37.687811	(11*s*, 6*p*)	−74.804559	−74.804558
(15*s*, 7*p*)	−37.688380	−37.688380	(13*s*, 7*p*)	−74.808117	−74.808117
(17*s*, 8*p*)		−37.688541	(15*s*, 8*p*)		−74.808945
(19*s*, 9*p*)		−37.688592	(17*s*, 9*p*)		−74.809266
(21*s*, 10*p*)		−37.688610	(19*s*, 10*p*)		−74.809350
(23*s*, 11*p*)	−37.688614	−37.688614	(21*s*, 11*p*)		−74.809381

	Sulfur			Selenium	
Basis	Optimal	Str. line	Basis	Optimal	Str. line
(10*s*, 6*p*)	−397.26057	−397.25823	(8*s*, 6*p*, 1*d*)	−2385.3307	
(12*s*, 7*p*)	−397.43035	−397.43023	(12*s*, 9*p*, 2*d*)	−2398.1209	
(14*s*, 8*p*)	−397.47966	−397.47966	(16*s*, 12*p*, 4*d*)	−2399.5044	
(16*s*, 9*p*)		−397.49538			
(18*s*, 10*p*)		−397.50121			
(20*s*, 11*p*)		−397.50352			
(22*s*, 12*p*)		−397.50431			
(24*s*, 13*p*)		−397.50462			

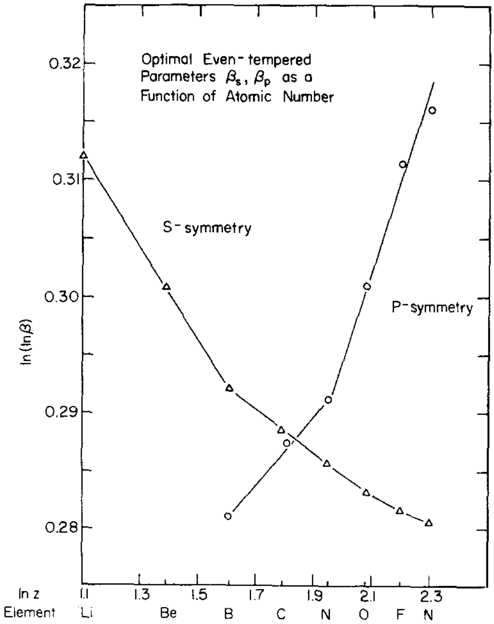


Fig. 4. Dependence of β_l for a (7*s*, 4*p*) basis on the atomic number

the periodic table, reflecting the fact that the built-in advantage of the latter in describing the cusp is becoming less important to the total energy.

Rapid convergence of the two columns results from the decreasing deviation of the optimal (α, β) points from the linearly interpolated values and from the simultaneous increase in the flatness of the energy surface as a function of the ET parameters. Additional optimization of the (23*s*, 11*p*) basis for carbon confirmed the validity of the "straight line" values of α_l and β_l at large N values.

Not only is the set of optimal ET parameters a smooth function of the number of Gaussians used for the expansion, it also behaves smoothly in going across the periodic table. In Fig. 4 the double logarithm of the optimal beta values for a (7*s*, 4*p*) ETG basis, as determined by Raffenetti [9], are plotted against the logarithm of the atomic numbers. Discontinuities in the slope are visible for both curves. For the *s* curve it occurs in going from Be to B. For the *p* curve it occurs between N and O.

4. Regularity of the Total Energy for Atoms

As the approximate SCF orbitals approach the integral transform representation of the exact atomic orbitals by means of the systematic sequence of (α, β) points given by Eqs. (13) and (14), the total energy approaches the HF limit in a very regular fashion. This can be seen in Figs. 5a and 5b where the beryllium energies of Schmidt and Ruedenberg [10] are used. Here the logarithm of the difference between each energy value and the near HF value of the (28*s*) basis is plotted as

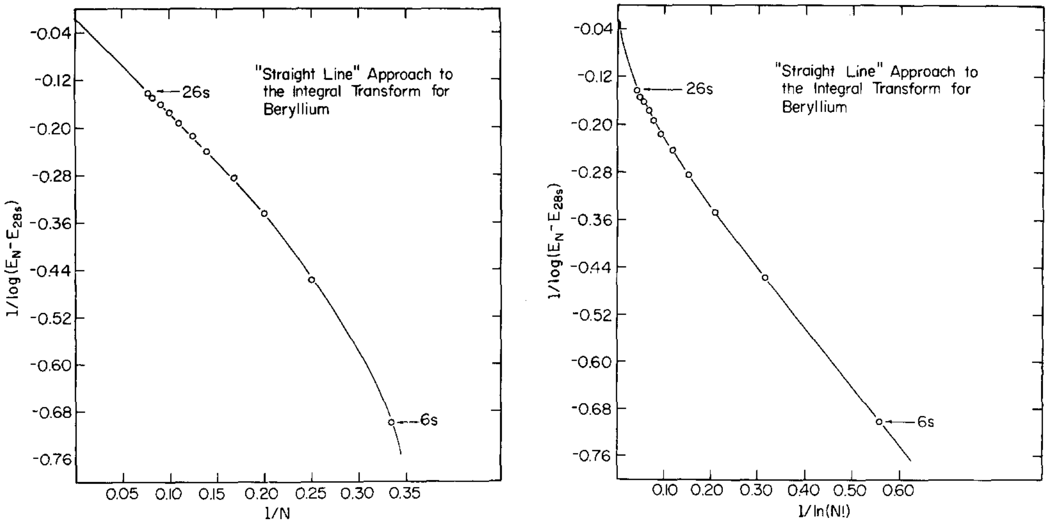


Fig. 5(a) Dependence of the error in the total energy as an inverse function of the number of basis functions. (b) Dependence of the error in the total energy as an inverse function of the factorial of the number of basis functions a function of $1/N$ and $1/\ln(N!)$ where N is half the number of s-primitives in the basis.

Because two different linear dependencies seem to dominate at opposite ends of the basis size spectrum a simple analytic expression combining the two such as

$$\operatorname{Log}(E_N - E_\infty) = AN/(1 + \exp[-\alpha(N-2)]) + B\ln(N!)/(1 + \exp[\alpha(N-2)]) \tag{15}$$

is capable of fitting the entire curve quite well. The constants A and B are determined linearly while E_∞ and α are determined nonlinearly by minimizing the standard deviation of the fit. The constant 2 appearing in (15) may also be varied, however, E_∞ is quite insensitive to its value. The nine Be points were fitted with a maximum deviation of 0.0002 and a typical deviation of 10^{-6} for the larger basis set points with $E_\infty = -14.573023$, $A = -0.551292$, $B = -0.664424$ and alpha $= 0.25$. Although this technique provides an extrapolation to the HF limit from a limited segment of the converging curve, its usefulness is limited to an improvement no better than one order of magnitude beyond the last computed value employed in the fit. For atoms, this degree of improvement is not too difficult to obtain by merely performing the indicated calculations with a larger basis set. The novelty of Eq. (15) lies in being able to accurately fit energies from such a wide range of basis sets with a two-term expansion, and with the help of such a fit, to predict with confidence, lower bounds as well as upper bounds for the exact limiting values. This will be elaborated on in a forthcoming paper by Schmidt and Ruedenberg [10].

Our basis sets are sufficiently large and the energy values sufficiently regular that with the use of Eq. (15) we can accurately estimate the HF limit for three of the atoms investigated. In applying Eq. (15) we identify N with $N(p)$, and note that $N(s) = 2N(p) + 1$ in carbon and oxygen while $N(s) = 2N(p) - 2$ in sulfur. The limits are: carbon $-37.688617\ E_h$, oxygen $-74.809397\ E_h$, and sulfur $-397.50488\ E_h$. The uncertainty in these values is ± 2 in the last digit. These values for the HF limit are in disagreement with the 1968 numerical Hartree–Fock energies of Fischer [11] by 0.00019, 0.00024, and 0.00122 for C, O, and S respectively, all lying above the Fischer values. Subsequent numerical calculations [12] are much closer to our estimates of the lower bound.

5. Optimal ET Molecular Parameters from Uncontracted Calculations

Since the reason for choosing a Gaussian primitive basis as opposed to a set of exponential primitives is the speed advantage the former gives in molecular multicenter integral evaluation, a pertinent aspect of the ET choice is that optimal ET molecular exponents are derivable from the atomic (α, β) with relative ease. Uncontracted optimizations of the ET (α, β) pairs in the carbon monoxide, methane and acetylene molecules with Gaussian basis sets of $(6s, 3p)$ up to $(14s, 7p)$ demonstrated that the energy differences between the optimal atomic and molecular (α, β) values for large sets were generally less than a millihartree in size. Moreover the optimal (α, β) values for the s-type primitives were very nearly identical for the atom and the molecule after 11*s*. To a large extent this is so because the majority of primitives for this symmetry are needed to refine the cusp. As these large exponent functions become an increasing percentage of the basis set, the (α, β) values which are optimal for the atomic cusp tend to dominate. This domination is aided by the near saturation of the valence region with sufficient functions such that exponent values can deviate considerably from the optimal values without substantial effect on the total energy. Optimal (α, β) pairs for *p*-symmetry in molecules are not observed to converge to the optimal atomic values as rapidly as the *s*-symmetry, but they too lie within a small region of the atomic values.

Accurate basis sets at the *atomic* limit are only a minimal requirement for accuracy in molecules. Uncontracted calculations on CO with optimal atomic even-tempered exponents show that there exists an "additional molecular error" beyond that which would be expected from the sum of the atomic errors seen in Table 5. This is illustrated in Table 6. Here the "additional molecular error" is defined as the difference between the error due to basis set truncation within each orbital symmetry for the *molecule* and the error due to basis set truncation within each orbital symmetry for the *two atoms*. It is seen that the error generally decreases with increasing basis set size. The magnitude of the molecular error also depends on the particular elements involved and the internuclear separations. The HF limit for CO comes from a hartree extrapolation of the total energies resulting from the largest three bases.

Table 6. Molecular errors for carbon monoxide

Basis	Energy (hartree)	Total molecular error (millihartree)	Addit. molecular error (millihartree)
a) Optimal atomic even-tempered parameters			
4*s*, 2*p*	−110.49971	2217.0	17.8
6*s*, 3*p*	−112.36763	349.1	5.1
8*s*, 4*p*	−112.64166	75.0	3.5
10*s*, 5*p*	−112.69584	20.9	3.8
12*s*, 6*p*	−112.71028	6.4	1.8
14*s*, 7*p*	−112.71478	1.9	0.8
16*s*, 8*p*	−112.71579	1.0	0.5
sp-HF limit	−112.7167	0.00	0.00
b) Optimal molecular even-tempered parameters			
4*s*, 2*p*	−111.02170	1695.0	527.7
6*s*, 3*p*	−112.39190	324.8	−24.7
8*s*, 4*p*	−112.64635	70.4	−4.9
10*s*, 5*p*	−112.69801	18.7	−2.5
12*s*, 6*p*	−112.71176	4.9	−1.5
14*s*, 7*p*	−112.71556	1.1	−0.8
16*s*, 8*p*	−112.71600	0.7	−0.3
sp-HF limit	−112.7167	0.00	0.00

Even with the simplifications inherent in the even-tempered approach, molecular optimization is still very time consuming. Moreover, while the set of optimal *atomic* parameters can be easily and accurately predicted for any size basis, we did not find this to be the case for the *molecular* set of small to medium size bases. Because of the similarity between atomic and molecular (α, β) pairs it is therefore reasonable to use the *atomic* even-tempered parameters in molecular calculations, and this choice is adopted in the further investigation.

6. Effective Contracted Orbitals for *s* and *p* Symmetries

Having determined the size of the primitive basis set and having adopted the atomic α, β parameters, we finally must choose a suitable set of contracted orbitals. To this end, we used a technique developed by Ruedenberg *et al.* [13] to *derive from an uncontracted molecular calculation that set of contracted orbitals which optimally reproduces the uncontracted result.* We then investigated the overlap of the space spanned by these *optimal* contracted orbitals with the space spanned by the occupied and virtual *atomic* SCF AO's. Our investigation involved ET basis sets ranging in size from (6*s*, 3*p*) to (14*s*, 7*p*) on the molecules CO, CH_4, C_2H_2, and H_2CO. In all cases studied the space of the 1*s*, 2*s*, 2*p* plus the first several virtual SCF AO's overlapped the space of the most important *optimal* contracted orbitals to better than 0.995 as shown in Table 7. It follows then that the space spanned by the occupied atomic SCF orbitals and the low-energy virtual atomic SCF orbitals furnish a near optimal set of contracted orbitals for molecular calculations.

The function space spanned by *these* contracted orbitals can be generated in a simpler manner however. Specifically, the *low-energy virtual* SCF AO's can be replaced by *the most diffuse primitive* AO's. The reason is as follows. Since the

Table 7. Transformation matrix between the HF atomic SCF AO'S (occupied plus first virtual) and the optimal contracted orbitals of CO

6*s* Basis				14*s* Basis			
Opt.	Atomic SCF			Opt.	Atomic SCF		
contr.	1*s*	2*s*	3*s*	contr.	1*s*	2*s*	3*s*
1	0.974	0.228	0.008	1	0.476	0.878	0.006
2	−0.229	0.974	0.011	2	0.879	−0.476	0.010
3	0.005	0.012	−0.991	3	0.009	0.043	−0.992

6*p* Basis			
Opt.	Atomic SCF		
contr.	2*p*	3*p*	4*p*
1	0.979	0.178	0.105
2	−0.196	0.631	0.750
3	0.015	−0.722	0.612

Table 8. Energy losses with respect to an uncontracted (16*s*, 8*p*) basis for a HF AO plus diffuse primitives set on CO

Contracted basis	Energy loss (millihartrees)
(6*s*, 5*p*)	0.4
(5*s*, 4*p*)	2.4
(4*s*, 3*p*)	5.8
(3*s*, 2*p*)	32.0
(2*s*, 1*p*)	149.0

virtual orbitals tend to span the continuum of the atomic HF eigenvalue problem, the energies of these orbitals increase monotonically with their kinetic energy, the lowest-energy virtual orbitals having the lowest kinetic energies. On the other hand, since the most diffuse primitives have the smallest kinetic and potential energies of all primitives, it turns out that *the lowest-energy virtual* SCF AO's *essentially consist of the most diffuse primitives orthogonalized to the occupied* SCF AO's. This orthogonalization is unnecessary, however, if the object is merely to span the same space. In fact, Raffenetti [14] was the first to compare the energies obtained with a set of HF AO's plus a diffuse primitive to Dunning's [15] contracted orbitals on the nitrogen molecule and water. He found them to be slightly better than Dunning's.

Because of the ease of generating this set and because of their similarity to the HF AO virtual space we shall employ this type of basis in the rest of this paper. Table 8 lists the energy losses incurred with such contractions for a (16*s*, 8*p*) primitive basis on CO. The quality of this contraction scheme depends slightly on the internuclear distances involved since the exact separated atoms' coefficients are built in.

We shall refer to basis sets generated in this manner as *Raffenetti-type* or SCFD bases. Specifically for an atom from Li to Ne, the symbol ETG16,8/4,3SCFD *basis* would denote a basis constructed from an atom-optimized even-tempered (16*s*, 8*p*) set of Gaussian primitives by contraction to a (4*s*, 3*p*) set, obtained by choosing the 1*s*, 2*s*, 2*p* atomic SCF AO's and the two most diffuse *s*-primitives and the two most diffuse *p*-primitives. It should be noted that such a contraction scheme is *not* of the *segmented* type. The calculation of molecular integrals between such basis orbitals from integrals between primitive functions requires therefore a procedure such as BIGGMOLI [14] in order to avoid computational inefficiencies due to duplication.

7. A Minimal Basis Set Function for Hydrogen

The hydrogen atom basis deserves special attention not only because of its ubiquitous appearance throughout chemistry but, more importantly, because of the substantial energetic effect which results from scaling its minimal basis function. Even though several contraction schemes now in use provide results within a fraction of a millihartree when two or more basis functions are used, it is nevertheless of interest to know which contracted function is most effective when used as a single minimal basis AO.

A common practice is to take its coefficients from the atomic 1*s* orbital and then determine the optimal scaling factor from the hydrogen molecule. A somewhat better single function is obtained by preserving the primitive even-tempered exponents from the isolated atom and taking as contraction coefficients those which yield the H_2 molecular orbital resulting from an uncontracted SCF calculation. Also when used in other molecules, this minimal basis function yields a lower energy error per H atom than the scaled 1*s* AO. This is illustrated in Table 9 which lists the errors for calculations on methane, H_2 and acetylene made with various hydrogen minimal basis set orbitals contracted from six *s*-primitives, with respect to calculations made with the uncontracted hydrogen 6*s* basis. In addition to the two contracted orbitals just mentioned the table also lists some results using minimal basis sets that yield optimal results in C_2H_2 and CH_4. The carbon basis in these calculations is a (6*s*, 3*p*) basis. The degree of contraction of the carbon basis influences the errors in Table 9 to less than 0.1 millihartrees.

Since acetylene and methane are usually found at opposite ends of the scaling range, the errors in other hydrocarbons are presumably no larger. For use as a single basis

Table 9. Errors per H atom in various molecules for hydrogen *minimal-basis-set-calculations* with respect to *uncontracted* hydrogen 6*s*-primitive-basis calculations[a]

Scale Factor	H_2	C_2H_2	CH_4
1. Coefficients of H minimal basis AO from atomic 1*s* AO			
1.00	18.7	24.9	20.9
1.19	0.1	4.5	3.6
1.29	4.9	2.6	1.8
1.35	13.0	1.7	2.4
2. Coefficients of H minimal basis AO from uncontracted H_2 SCF MO			
1.00	0.0	4.1	3.1
1.08	5.0	1.3	1.1
1.10	7.0	1.0	1.1
1.13	11.9	0.9	1.2
1.15	17.2	1.2	1.5
3. Coefficients of H minimal basis AO from uncontracted SCF calculations on C_2H_2			
1.00		0.5	1.9
4. Coefficients of H minimal basis AO from uncontracted SCF calculations on CH_4			
1.00		2.2	0.4

[a] All errors are given in millihartrees.

Table 10. Atomic and molecular hydrogen errors with the optimal H_2 contracted orbital and diffuse primitives for a 6*s* basis[a]

Scale factor	Contracted basis	Atomic	Molecular
1.00	one *s* orbital	18.6	0.4
1.00	two *s* orbitals	1.9	0.4
1.00	three *s* orbitals	0.2	0.4
1.00	uncontracted basis	0.2	0.4
1.10	one *s* orbital	48.5	14.4
1.10	two *s* orbitals	2.7	1.7
1.10	three *s* orbitals	0.3	0.5
1.10	uncontracted basis	0.3	0.5

[a] All errors are in millihartrees.

function the minimal basis No. 2 with a scaling factor of 1.08 would seem to represent an optimal compromise.

In cases where a hydrogen atom will dissociate from the molecule or where additional accuracy is required, some diffuse primitives must be added to increase the flexibility of the basis. The quality of various basis sets of this type is illustrated in Table 10. The *coefficients* of the minimal basis orbital are those from Case 2 of Table 9, (i.e. from an uncontracted H_2 calculation with scale factor unity). The second and third orbitals, where present, are the one or two most diffuse single Gaussian primitives. The error listed is with respect to the exact value in the atom and with respect to the *s*-limit of the SCF approximation in the H_2 molecule.

It is apparent that, when at least two *s* orbitals are used, the choice of a scale factor of unity will give equally satisfactory results, within a millihartree, for the free H atom as well as for the H orbital in a molecule. From the data given in Table 9 for Case 2, it can be inferred that this choice will also give millihartree accuracy for hydrogen in other molecules.

8. Polarization Functions

In order to construct symmetry orbitals for use in molecular calculations, admixtures of all primitive functions (or combinations of primitives transforming according to the irreducible representations of the molecule's point group) should be included in the algorithm. While for atoms in the first and second rows this restricts the primitives to be of *s* or *p* symmetry, functions of higher angular momentum may mix in for molecules. Such functions allow the MO's to polarize in the direction of the bond and were initially suggested by Nesbet [16]. Polarization functions are known to provide a substantial energy lowering and improvement of expectation values when compared with similar calculations without such functions. In the CO molecule, approximately three or four sets of even-tempered *d* primitives are required in order to attain millihartree deviations from the *s*, *p*, *d* basis set limit.

Table 11. Optimal ET *d*-symmetry parameters and energies for CO

Basis[a]	Carbon Alpha	Beta	Oxygen Alpha	Beta	Energy
(10*s*, 5*p*/5*s*, 3*p*)	–	–	–	–	−112.6958
(10*s*, 5*p*, 1*d*/5*s*, 3*p*, 1*d*)	1.00000	1.09660	1.00000	1.03890	−112.7619
(10*s*, 5*p*, 2*d*/5*s*, 3*p*, 2*d*)	0.06588	3.89619	0.03088	5.87085	−112.7680
(10*s*, 5*p*, 3*d*/5*s*, 3*p*, 3*d*)	0.06856	3.81904	0.04689	4.22070	−112.7704
(10*s*, 5*p*, 4*d*/5*s*, 3*p*, 4*d*)	0.06050	3.62110	0.02814	4.11723	−112.7712

[a] Contracted orbitals are the optimal contracted orbitals for 10*s*, 5*p*.

Table 11 shows the energy lowering with the inclusion of *d* functions optimized for the CO molecule. The 1*d* and 2*d* exponents were optimized with a (6*s*, 3*p*) and (8*s*, 4*p*) respectively, instead of with the (10*s*, 5*p*) basis. Although a satisfactory description of some properties may be obtained without such functions, others, like the internal rotation in hydrogen peroxide, require that they be present in the basis set. A recent study by Poirier and Kari [17] indicates that for the computed one-electron properties of first- and second-row hydrides there is no economic justification for including *f*-symmetry polarization functions. For CO they lower the total energy by at most 6 millihartrees, since the actual restricted HF limit for CO lies within a couple millihartrees of $-112.7892\ E_h$, obtained with a very large *s, p, d, f* Slater-type orbital (STO) basis [18].

We shall use the notation ETG16,8/4,3SCFD+2,1 to indicate the addition of two *d* primitives and one *f* primitive as polarization functions to the basis defined at the end of Sect. 7.

9. Regularity of the Total Energy and Dipole Moment for CO

In the case of the free atom a systematic approach to the complete basis provided sufficient regularity in the total energies that extrapolation to the integral transform limit became feasible. Similar behavior is found in the carbon monoxide molecule.

Table 12. Hartree extrapolations of CO total energies

Basis	Energy	*E* (limit)
(6*s*, 3*p*)	−112.3919	–
(8*s*, 4*p*)	−112.6464	–
(10*s*, 5*p*)	−112.6980	−112.711
(12*s*, 6*p*)	−112.7118	−112.717
(14*s*, 7*p*)	−112.7156	−112.717
(16*s*, 8*p*)	−112.7160	−112.716
(6*s*, 3*p*, 1*d*)	−112.4919	–
(8*s*, 4*p*, 2*d*)	−112.7231	–
(10*s*, 5*p*, 3*d*)	−112.7704	−112.783
(12*s*, 6*p*, 4*d*)	−112.7815	−112.785

Table 13. Dipole moments from various ET basis sets on CO[a]

Basis *A*		Basis *B*	Basis *C*		Basis *D*
(6*s*, 3*p*)	−112.362	−112.426	(6*s*, 3*p*, 1*d*)	−112.491	–
	0.641	0.581		0.088	–
(8*s*, 4*p*)	−112.640	−112.649	(8*s*, 4*p*, 2*d*)	−112.723	–
	0.552	0.567		0.336	–
(10*s*, 5*p*)	−112.696	−112.701	(10*s*, 5*p*, 1*d*)	−112.770	–
	0.506	0.543		0.212	0.258
(12*s*, 6*p*)	−112.7092	–		–	–
	0.479	–		–	–

[a] Every entry contains the total energy (in hartrees) in the first row and the dipole moment (in debyes) in the second row. Basis set *B* contains the functions in Basis *A* plus some additional diffuse primitives. Basis *D* consists of the functions in Basis *C* plus some additional diffuse primitives.

However, the increase in basis set size in going from an atom to a molecule precludes the use of as large a set as was used in the atoms. Table 12 lists the results of Hartree extrapolations on the total energies obtained with *s*, *p*, and *s*, *p*, *d* basis sets.

Another important molecular property is the dipole moment. Since this property is rather sensitive to the basis set's ability to properly span a region of space other than near the nucleus it is of interest to see whether the use of an energy optimized ET basis allows an extrapolation of the values obtained with smaller bases. In Table 13 the values of the dipole moment from polarized and nonpolarized ET bases are reported. In some cases additional diffuse primitives with an *s* exponent of 0.06 and a *p* exponent of 0.03 were added to the basis to help in describing the region of space far from the nucleus.

The HF limit value is close to 0.276 D and the experimental value obtained by microwave spectroscopy is −0.112 D. The fact that the HF value has incorrect sign is not of concern to us for this work. What seems evident from these results is that the value of this property is too highly dependent on diffuse primitives in the basis to allow extrapolation. Even though basis sets *A* and *B* or *C* and *D* must converge to the same limit they are still far enough apart in their values that a simple extrapolation would seem difficult.

Acknowledgements. The authors wish to thank Mr. M. W. Schmidt for his permission to use data reproduced in Fig. 5 from work which represents a further extension of the present study [10].

The present work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences.

References

1. Schwartz, C., in: Methods in computational physics, Alder, B., ed., Vol. 2, p. 241. New York: Academic Press 1963
2. Ruedenberg, K., Raffenetti, R. C., Bardo, R. D., in: "Energy Structure and Reactivity" Proceedings of the 1972 Boulder Conference on Theoretical Chemistry, p. 164. New York: Wiley 1973

3. Huzinaga, S.: Approximate atomic functions. University of Alberta Press 1971
4. Kikuchi, R.: J. Chem. Phys. **22**, 148 (1954)
5. Bishop, D. M., Somorjai, R. L.: J. Math. Phys. **11**, 1150 (1970)
6. Taylor, W. J.: J. Math. Phys. **19**, 52 (1978)
7. Raffenetti, R. C.: J. Chem. Phys. **59**, (1972)
8. Bardo, R. D., Ruedenberg, K.: J. Chem. Phys. **59**, 5956 (1973)
9. Raffenetti, R. C.: Intern. J. Quantum Chem. **9**, 289 (1975)
10. Schmidt, M. W., Ruedenberg, K.: to be published
11. Fischer, C. F., in: Some Hartree-Fock results for atoms helium to radon. Vancouver, B.C.: University of British Columbia Press 1968; Atomic Data **4**, 301 (1972)
12. Fischer, C. F.: The Hartree-Fock method for atoms: a numerical approach. New York: Interscience 1977
13. Bardo, R. D., Ruedenberg, K.: J. Chem. Phys. **60**, 918 (1974); Cheung, L. M.: Ph.D. dissertation, Appendix B, Department of Chemistry, Iowa State University, 1975
14. Raffenetti, R. C.: J. Chem. Phys. **58**, 4452 (1973)
15. Dunning, T. H.: J. Chem. Phys. **53**, 2823 (1970)
16. Nesbet, R. K.: J. Chem. Phys. **36**, 1518 (1962)
17. Poirier, R., Kari, R.: Can. J. Chem. **56**, 543 (1978)
18. Green, S.: J. Chem. Phys. **52**, 3100 (1970)

Received March 13, 1979