Reviews in Computational Chemistry, Volume 1
Edited by Kenny B. Lipkowitz, Donald B. Boyd

CHAPTER 1

Basis Sets for *Ab Initio* Molecular Orbital Calculations and Intermolecular Interactions

David Feller and Ernest R. Davidson

Chemistry Department, Indiana University, Bloomington, Indiana 47405

INTRODUCTION

Of the three primary computational chemistry tools, two (molecular mechanics and semiempirical molecular orbital theory) rely on embedded empirical parameters, while the third (*ab initio* quantum chemistry) is potentially capable of reproducing experiment without such parameters. Since the first two methods depend on the availability of reliable experimental data, they are best applied in situations requiring an interpolation between known experimental quantities. The quality of the fundamental parameters on which these models are based depends on the quality and quantity of experimental data.

Within the realm of *ab initio* methods one should distinguish two different approaches. In the "calibrated" approach, favored by Pople and coworkers, the full exact equations of the *ab initio* method are used without approximation. The basis set is fixed in a semiempirical way, however, by calibrating calculations on a variety of molecules. The error in any new application of the method is estimated based on the average error obtained, compared with experimental data, on the calibrating molecules. This is different in philosophy from the "converged" approach favored by chemical physicists interested in small molecules. In the latter approach, a sequence of calculations with improving basis sets is done on one molecule until convergence is reached. The error in the calculation is estimated from the sensitivity of the result to further refinements in the basis set. Clearly the calibrated method is the only one that is practical for routine use in computational chemistry. Converged *ab initio* calculations are limited to a few small molecules and normally are done only when developing and testing new methods.

Ab initio theory more easily lends itself to situations in which little or no experimental information is available, i.e., in *new* areas of chemistry. However, to the extent that the *ab initio* approach is independent of experimental data, it finds itself at a disadvantage relative to other methods in the areas of computational speed and accuracy. Viewed from a broader perspective, each methodology is seen to have its own particular strengths and weaknesses and an intelligent use of these tools, therefore, implies some understanding of their inherent limitations.

Since the overwhelming majority of *ab initio* calculations ultimately expand a wavefunction in terms of Gaussian basis functions, this review is an attempt to provide the nonexpert with rough guidelines for selecting from among the many basis sets in the chemistry literature. In conjunction with the assumed form for the many-electron wavefunction, the choice of basis set represents one of the principal adjustable parameters in any *ab initio* calculation. A wrong choice can render the results of even large and time-consuming calculations meaningless. If competitive, fully numerical methods were suddenly to appear, no doubt the quantum chemistry field would see a rapid abandonment of the basis set approach to building wavefunctions. Although research continues along those lines,[1] no general numerical method for polyatomic molecules has appeared. Meanwhile, in spite of the fact that the topic of basis sets is often considered highly technical, at least a cursory understanding is essential to workers who must apply *ab initio* methods to questions of chemical interest.

The sheer number of basis sets, compounded by the ever-increasing number of properties that might be desired, makes the choice a nontrivial one. It is especially difficult for chemists who are primarily interested in molecular mechanics or semiempirical methods, but who find themselves either wanting to perform *ab initio* calculations or having to critique a colleague's use of *ab initio* methods. Even experts in the field find it difficult to keep up with the literature as the proliferation of new basis sets continues unabated.

The three primary computational chemistry methodologies are not mutually exclusive in their range of applicability. For example, all three are capable of predicting the structure of molecules and all three provide relative conformational energies. This should not be misconstrued to mean that they behave identically. For example, molecular mechanics[2] can often reproduce bond lengths to an accuracy of ±0.004 Å in molecules similar to those used in parameterizing the force field, while minimal level Hartree–Fock *ab initio* methods require significantly more computer resources to achieve results that may be an order of magnitude less accurate. Of course, more accurate *ab initio* treatments can match the accuracy of molecular mechanics, but at an increased cost in computer time. In fact, some molecular mechanics force fields are now being parameterized with *ab initio* results because of the dearth of experimental data for particular classes of compounds.

Similarly, semiempirical methods, such as MNDO,[3] which are parameterized to reproduce heats of formation, have been reported to yield ΔH_f's with an accuracy of ±9 kcal/mol[4] for a selected set of molecules. Simple *ab initio* calculations may be more than a factor of two less accurate and 5–10 times slower. Again, in most cases it is possible to perform more elaborate *ab initio* calculations that can reduce this error to ±2 kcal/mol or less.

Nevertheless, the combination of generality and flexibility makes *ab initio* methods a powerful complement to experimental measurements as well as to other computational techniques for small to intermediate size systems. *Ab initio* methods can, in principle, be applied to any geometry on the ground state or any excited state potential energy surface. Furthermore, with modern computer programs the "quality" of the wavefunction can be easily improved, in contrast to molecular force field or semiempirical methods that are comparatively difficult for the user to improve systematically.

In broad terms, molecular mechanics seeks to provide information about molecular structure and relative energies. In addition, some programs can provide vibrational normal mode information. Several commonly used semiempirical methods are parameterized on experimental heats of formation at 25°C and other properties and have been found to give reasonable geometries (with accuracies in the ±0.02 Å range) and selected one-electron properties, such as dipole moments (with accuracies of ±0.4 D).

Even though molecular mechanics and semiempirical methods are capable of describing a great many molecular properties, and more are being added every year, certain properties are beyond the scope of either approach. In such cases *ab initio* methods may offer the only theoretical model capable of attacking the problem. For example, neither method can accurately treat the manifold of valence and Rydberg excited states of linear polyenes. Nor can they handle the hyperfine spin properties of first- or second-row elements.

The most recent reviews of basis set technology were written three or more years ago.[5–8] Since then several new basis sets have appeared in the theoretical literature aimed at further minimizing the computational expense of *ab initio* work while still maintaining an acceptable level of agreement with experiment. The computer time required for some parts of an *ab initio* calculation is so strongly dependent on the number of basis functions that relatively minor changes in the basis can render a calculation intractable. At the Hartree–Fock level the numerous electron–electron repulsion integrals, which are necessary to solve the SCF equations, increase as the fourth power of the number of Gaussian functions. Commonly used post-Hartree–Fock methods, which strive to improve the wave function by correlating the motions of the electrons, vary in their computational expense anywhere from a modest n^5 to an imposing $n!$, where n is the number of basis functions.

SOME TERMINOLOGY

Gaussian Compared to Exponential Functions

The use of Cartesian Gaussian-type orbitals (GTOs) in *ab initio* work may come as a surprise to anyone who recalls the functional form of the hydrogen atom orbitals, $\chi_{\text{hydrogen}} \sim \exp(-\zeta r)$. Cartesian GTO's have the form:

$$\chi_{\text{GTO}} = N x^l y^m z^n \exp(-\zeta r^2)$$

where N is a normalization constant, ζ is a constant called the "orbital exponent," and distances are measured with respect to the position of the Gaussian in three-dimensional space. Customarily one defines $L = l + m + n$, and refers to $L = 0$ functions as "s" functions, $L = 1$ as "p" functions, $L = 2$ as "d," etc. A Gaussian function clearly has the wrong behavior near the origin (zero slope instead of a cusp) and at large r, i.e., $\exp(-r^2)$ drop-off instead of the slower $\exp(-r)$. Nodeless Slater-type orbitals (STOs)[9] of the form

$$\chi_{\text{STO}} = N r^{n-1} \exp(-\zeta r) Y_{lm}(\Theta,\phi)$$

where Y_{lm} is the usual spherical harmonic, would seem to be a much better match for the hydrogenic orbitals and, in fact, are still the functions of choice in atomic or diatomic calculations.[10] Although much has been written about the intrinsic deficiencies of Gaussians relative to exponential functions, the reason for adopting the former over the latter was purely pragmatic. Boys[11] first advocated the use of Gaussians in polyatomic molecules because they made the four-center, two-electron integrals relatively easy to compute. Although three to four times as many s- and p-type Gaussians might be needed on first-row elements B–F, compared to exponential functions, computer programs to handle Gaussians were relatively easy to write and, more importantly, they could be made fast enough to more than offset the larger number of functions. The advantages enjoyed by exponential functions is even less for functions with higher L values. Here a single STO is worth less than two GTOs. Utilization of concepts developed in the 1960s, which will be described in the following section, further diminishes the advantage of exponential functions over Gaussians.

Contracted Gaussians

The term "atomic orbital" (AO) often appears in the early quantum chemistry literature without the benefit of a precise definition. The expression was loosely used in the context of combining functions derived from the free atom in such a way as to construct an approximation to the molecular orbital, i.e., the linear combination of atomic orbitals (LCAO) approximation. STOs were introduced as an approximation to atomic orbitals. In this review we shall use the term "atomic orbital" to mean a Hartree–Fock (HF) orbital obtained from an atomic calculation. As a consequence of the AO basis set concept, a "minimal" basis set would contain only those functions whose counterparts were occupied in the atoms from which the molecule was constructed. For example, a minimal basis on methane would consist of 1s, 2s, and 2p functions for carbon and four 1s functions, one for each hydrogen.

Work by Frost and co-workers[12] in the mid 1960s abandoned the idea of AO-based functions to arrive at an even more compact basis set. They suggested the use of s-type Gaussians that were not fixed at the atomic centers, but could "float" in space so as to optimally represent each localized pair of electrons. Because only one function was needed for each *pair* of electrons, the basis sets used in "floating spherical Gaussian" (FSGO) scheme were often referred to as "subminimal." Extensions of Frost's idea to ellipsoidal Gaussians[13] of the form

$$\chi_{\mathrm{EGTO}} = N \exp(-\{\alpha x^2 + \beta y^2 + \gamma z^2\})$$

where N is a normalization constant and α, β, and γ are now permitted to assume different values, improved the quality of the wave function without increasing the number of basis functions. However, the idea of using a single Gaussian, whether spherical or ellipsoidal, to represent an electron pair has fallen out of favor as *ab initio* methods have been pushed to greater accuracy.

A compromise between the AO basis set approach and floating spherical Gaussian approaches was proposed by Whitten[14] in 1966. He suggested the use of "Gaussian lobe" functions, which were s-type Gaussians situated about the nucleus in such a way as to mimic functions of higher angular momentum. As with floating spherical Gaussians, the use of Gaussian lobe functions has greatly decreased because it was difficult to systematically improve the basis set. Problems with the loss of numerical precision appeared as higher angular momentum functions, such as f- and g-type functions, were approximated by combinations of lobe functions. At the same time, very efficient programs that could handle Cartesian Gaussians up through g functions began to appear. Almost all *ab initio* calculations are now done with atom-centered Cartesian Gaussians.

By the early 1960s, as more experience was gained in performing *ab initio* calculations, it was realized that fixed linear combinations of Gaussians, known as "contracted" functions,[15,16] could be used in place of the individual, or "primitive," functions with an acceptably small decrease in quality of the wave function. This development was based on the recognition that most of the s-type Gaussian primitives were present in the basis set to help describe the energetically important, inner shell, region of space near the nucleus. Since the atomic inner shells do not experience much change when a molecule is formed, it was possible to determine the coefficients at the atomic level and simply transfer them to the molecular situation.

The shape of the inner portion of the 2s orbital is largely determined by the nuclear potential and by orthogonality constraints imposed by the 1s function. Therefore, it will also have contraction coefficients for the primitive Gaussians in the inner core region (large orbital exponents), which are transferable between the free atom and molecules. The approximation imposed by deriving contraction coefficients for inner shell type Gaussians from atomic calculations is adequate as long as the 1s or "core" electrons do not actively participate in the chemical process under examination. Conveniently, this is the case for the overwhelming majority of uses to which *ab initio* calculations are now being put.

Even p functions, which have a node at the nucleus, have their shapes in the energetically important small r region largely determined by the partial shielding of the 1s orbital. As a consequence, the coefficients for tight p functions (those with large orbital exponents) change very little when the atom finds itself part of a molecule. Thus, although contraction coefficients have been derived in a wide variety of ways, most are obtained from free atom calculations. The implicit assumption is that the coefficients determined for a particular atom can be used in any molecule containing that atom.

In fact, in a contraction scheme first proposed by Raffenetti,[17] the minimal basis set functions are taken to be the actual HF AOs, obtained by first performing uncontracted calculations on the free atoms. For a variety of molecules, it has been demonstrated that these functions overlap the space of the "optimal"[18] molecular minimal basis set to better than 0.99, thus, empirically justifying the assumption mentioned in the preceding paragraph.

The "art" in designing basis sets comes in the treatment of the valence regions of space, where molecular formation causes large distortions from a free atom environment. Care must be taken not to impair the flexibility of the basis set in this critical region, while simultaneously keeping the total number of functions as small as possible. The next step up from minimal atomic orbital basis sets is, then, to split the description of the 2s and 2p orbitals, which is handled by a single 2s and set of 2p orbitals in the minimal basis set approximation, into two functions each. This results in what is known as a "split valence" (SV) basis. For investigations that do not require high accuracy, SV basis sets are probably the most frequently used. A comparison of typical results from various SV basis sets and experiment will be discussed subsequently.

Slightly more flexible than the split valence basis set is what is known as the "double zeta" (DZ) basis. For first-row atoms (Li–Ne) a DZ basis will have four functions of s symmetry (compared to only three for an SV basis) and two functions of p type. The label "double zeta" derives from the time when most calculations were being done on diatomics with STO basis sets.

The exponents were often denoted by the Greek letter zeta. Thus, a "single-zeta" basis set would have a single exponential function representing each atomic orbital. A double-zeta basis would have two exponential functions for each AO. The terminology has carried over into Gaussians.

Polarization Functions

Basis functions with higher L values may also be added to the expansion set to better account for the distortion from atomic symmetry that results from placing the atom in a molecular environment. These functions usually go by the name "polarization" functions because they permit the polarization of the AOs. A double-zeta basis, when augmented with such functions, is called a "double-zeta plus polarization" (DZP) basis set.

This style of jargon can be (and has been) carried further, giving rise to phrases such as "triple-zeta plus polarization" (TZP) or "triple-zeta plus double polarization" (TZDP). However, the terminology for basis sets beyond the DZP level rapidly becomes so unwieldy that a kind of shorthand is commonly invoked. The uncontracted basis set is given in parentheses, followed by the contracted basis in square brackets. If the molecules of interest contain atoms from different periods of the periodic table, their basis functions are separated by a slash. For example, a double-zeta contracted basis set on water, which likewise contained p functions on hydrogen and d functions on oxygen, might be denoted as (10s,5p,1d/5s,1p) → [4s,2p,1d/2s,1p], where the original Gaussian primitive basis set contained 10 s functions and 5 p functions on oxygen and 5 s functions on hydrogen. Some additional textual information about the manner in which the primitives were contracted is also necessary before the basis set can be completely specified. Pople and co-workers use a different notation, which is slightly more informative. For example, their 6-31G* basis is a SVP basis (10s,4p,1d) → [3s,2p,1d] with six Gaussians combined to form the first s contraction, three combined to form the second s combination, and the third s being an uncontracted primitive. Similarly, three primitives are combined to form the first contracted p, while the second is uncontracted.

Another subtlety concerns the set of higher L Cartesian Gaussians, i.e., d, f, and g functions. There are six Cartesian d's with $l + m + n = 2$. The $d_{xx} + d_{yy} + d_{zz}$ combination of these corresponds to a function of atomic s symmetry. Sometimes this combination is included in calculations and sometimes it is omitted. When one compares literature results from various sources, it is important to know whether the s component of the d's was present. With many commonly used basis sets the effect, even with only one d set, is not negligible. When adding a set of d functions to the 6-31G basis,[19,20] for example, the difference in the SCF energy between keeping or omitting the s component of the d set is on the order of 1 millihartree (0.6 kcal/mol) for each first-row atom present in the molecule. Unfortunately, there is no standard notation that tells whether the s component of the d's has been kept. For example, the Pople basis sets 3-21G$^{(*)}$ and 6-31G* both include the s component while 6-311G* does not.

The replacement of the Gaussian primitive basis with a smaller number of contracted basis functions drastically reduces the amount of storage required to hold the two-electron repulsion integrals as well as the time required for the self consistent field (SCF) and electron correlation portions of the calculation. However, the time required to compute the two-electron integrals, or for that matter any set of integrals over the basis functions (e.g., gradient integrals used in geometry optimizations), is not reduced by the use of contracted functions.

Comparisons between STOs and contracted Gaussian basis sets in water,[21] copper,[22] and other systems[23] showed comparable accuracies with comparably sized basis sets.

Complete Sets

At this point the reader may be wondering where it all ends. In theory, the answer is never. To construct a "complete" basis set, capable of exactly representing the Hartree–Fock wave function for any molecule, it would be necessary to include an infinite number of functions of each symmetry type (s,p,d,f, . . .). This is sometimes referred to as the "Hartree–Fock limit." For an in-depth examination of this issue the reader is referred to representative work by McDowell[24] and Klahn.[25] Although a rigorous examination of completeness is beyond the scope of the present treatment, it is helpful to consider a more practical definition of completeness that allows for real world limitations. We thus arrive at the notion of *effectively* complete basis sets.

A working definition of "effective completeness" might vary somewhat depending on a person's area of interest, but most people would agree that a basis set capable of reproducing the Hartree–Fock limit energy to within 1 millihartree (1 hartree = 627.5 kcal/mol = 27.2 eV) along with other properties to within ±1% is nearly complete. To give one concrete example of what it takes to achieve this level of accuracy, the SCF energy of CO with a very large basis consisting of only s- and p-type functions is given in Table 1 as -112.717 hartree. Adding d functions until the d space is saturated lowers the energy by 0.068 hartree. Higher L functions (f, g, etc.) add another 0.006 hartree for a final energy of -112.791. So the effect of polarization functions on the energy is quite large, even for small molecules. Other properties can be equally affected. The SCF dipole moment of CO changes by nearly a factor of two (from 0.54 to 0.27 D) as higher L functions are added (but is still of the wrong sign because of electron correlation effects).

The effects produced by such functions are even more dramatic when correlated wave functions are used. The Hartree–Fock wavefunction for oxygen makes use of s- and p-type functions only. However, at the configuration interaction (CI) level, higher L functions can be used to correlate the motions of the electrons. Table 1 shows how the CI energy and the isotropic hyperfine

Table 1 Effects from Adding Higher L Functions at the SCF and CI Levels of Theory[a]

	CO	
	E(SCF)	μ (D)
(s,p) limit	−112.717	0.54
(s,p,d) limit	−112.785	0.26
(s,p,d,f) limit	−112.790	0.27
Numerical HF	−112.791	0.27
	O atom	
	E(CI)	A_{iso} (MHz)
(s,p) limit	−74.935	−26
(s,p,d) limit	−75.032	−30
(s,p,d,f) limit	−75.053	−30
(s,p,d,f,g) limit	−75.061	−31
Experimental	−75.069	−34.5

[a] Estimated limits based on large uncontracted even-tempered Gaussian calculations. The CI results for oxygen are based on the estimated full CI energy derived from large multireference single- and double-excitation calculations.

property, which is a measure of the unpaired spin density at the nucleus, converge as a function of L.

Experience has shown that it is presently possible to effectively converge basis sets for small molecules containing one to three first-row atoms. For larger molecules, which embody the majority of chemistry, smaller sets must be used. However, the accuracy from even these smaller sets is often sufficient to answer questions of immediate chemical interest. The available accuracy for every-sized molecular system will continue to improve as rapid advances in computer hardware and software are realized. Thus, although the long-term goal is to be able to compute quantities such as energy differences reliably to within ± 1 kcal/mol and bond lengths to ± 0.01 Å, in the short term there are problems faced by chemists where useful insights can be gained with more approximate results.

In summary, nearly all *ab initio* calculations currently use contracted Gaussian basis functions. Uncontracted sets have been relegated to investigations seeking very high accuracy in small systems where it might be desirable to retain as much flexibility as possible.[26]

The Basis Set Superposition Error

As *ab initio* calculations were pushed to greater accuracy and researchers began tackling problems such as the weak van der Waals interaction between rare gas atoms or the hydrogen bonding in water, it became apparent that the use of incomplete basis sets resulted in significant errors in computing potential energy curves. The origin of the so-called "basis set superposition error" (BSSE) comes from the tendency for electrons associated with an atom having a deficient basis set to make use of basis functions on neighboring atoms to lower their energy.

For example, each atom in a simple diatomic molecule, such as CO, may experience two sources of energy lowering as they approach each other. Not only will there be a lowering due to bond formation, but an artificial lowering may arise as a result of an improved description of each atom. Even in the absence of the other nucleus and its electrons, the energy of either carbon or oxygen by itself may show a decrease when its own basis is augmented by the collection of basis functions associated with the other center.

Several schemes have been proposed for estimating the BSSE. The most widely used is the counterpoise method,[27] which defines the BSSE for a dimer as the sum over both monomers of {*E*(monomer in monomer basis) − *E*(monomer in full dimer basis)}. Some authors caution against allowing the electrons on monomer A to utilize the function space corresponding to the occupied orbitals on monomer B.[28,29] More recently this controversy seems to have been settled in favor of the full counterpoise correction.[30,31] Although this effect causes severe problems for studies involving small energy changes, most computed one-electron properties, such as dipole moments, do not appear to be affected by it.[32]

CHOOSING A BASIS SET

In early *ab initio* programs specifying the basis set was a laborious and error-prone task. The user was required to punch several dozen floating point numbers into specific columns of the input deck without a single typing error. With current programs (e.g., GAUSSIAN,[33] HONDO,[34] and GAMESS[35]) the user may select from among a wide variety of internally defined basis sets by simply specifying the appropriate keyword (usually the basis set name). While this represents a significant step forward in making the programs easier to use, it is still not ideal. Rather than choosing from among alternatives like "STO-3G," "3-21G," and "DZP," which assume some familiarity with the jargon surrounding basis sets, most users would prefer to simply indicate the properties they want and the desired accuracy in the final answers. Unfortunately, theory is still unable to predict the accuracy of a given calculation on purely formalistic grounds. The best that can currently be offered is to calibrate the method (i.e., basis set and level of correlation recovery) against experiment or numerical Hartree–Fock methods and hope that the molecule of interest behaves similarly to the calibration set. Some of the difficulties in a priori error estimates are discussed in the recent work of Feller et al.[36] on small molecules.

Molecular Geometries

The most widely calibrated basis sets for general purpose use are those of Pople and co-workers, which range from the small STO-3G minimal basis all the way up to basis sets with extra diffuse functions and multiple sets of higher L functions. Besides being widely tested, these basis sets offer the computational advantage that, when used with computer programs that exploit their features, the calculation of the two-electron and first derivative integrals is faster than for most other basis sets because exponents are shared between s and p functions. Tavouktsoglou and Huzinaga[37] developed a basis similar to the STO-3G basis but, instead of determining the contraction coefficients by a least-squares fit to STOs, they used the atomic SCF coefficients.

Table 2 provides an indication of how well the Pople-style basis sets and some of the most widely used non-Pople basis sets succeed at predicting molecular geometries. For a more detailed examination of the Pople-style basis sets, the user is directed to the book by Hehre et al.[38] that provides a more comprehensive analysis of the magnitude of errors to be expected from various classes of compounds. It should be noted that comparisons of *ab initio* results with experiment are complicated by the fact that most theoretical calculations produce results appropriate to nonvibrating molecules at the true minimum in the multidimensional stretching and bending potential surface. Low temperature, gas phase experimental results are the closest comparable quantities, but are often unavailable. Because of this, Hehre et al. warn that differences between theoretical and experimental geometries of less than 0.01 Å or 1° are usually meaningless, except in very small molecules. Corrections to properties such as the dipole moment due to vibrational motion are usually less than 10% and are often neglected unless a very precise determination is desired.

The STO-3G basis was initially developed for the elements H–Ne.[39] However, over the years it was expanded to cover Na–Ar,[40] P–Kr,[41] and Rb–Xe.[42,43] A glance at Table 2 shows that, not surprisingly, the simple STO-3G basis, which represents each atomic orbital by a single contracted function designed to mimic a Slater-type orbital, shows the largest average error. Nevertheless, the average deviation in bond lengths (~0.03 Å) and bond angles (~2.1°) for first-row AH_n compounds is quite good considering the crudeness of the basis set. For slightly larger compounds, generically labeled $A_lB_mH_n$, the average errors are even smaller (±0.03 Å, ±2°).

In general, this basis does better at reproducing multiple bonds between first-row atoms than it does for single bonds. Carbon–carbon single bonds in unstrained systems are an exception to this rule. Compensating errors allow the STO-3G Hartree–Fock C—C bond lengths in compounds such as ethane and propane to come within 0.01 Å of the experimental values.

In spite of the reasonably small average errors just mentioned, caution must be exercised in the use of STO-3G basis. If a different list of compounds different from those chosen for Table 2 were selected for the comparison, quite

Table 2 Representative RHF and Correlated Results Obtained with Some Commonly Used Basis Sets

				Average errors in geometries and other properties[a]				
Name	Contracted primitives	Functions	Elements	AH_n	$A_lB_mH_n$	Dipole	BSSE	VIP
STO-3G[b]	(3s)	[1s]	(H–He)					
	(6s,3p)	[2s,1p]	(Li–Ne)	SCF: ±0.032 Å, ±5.6°	±0.027 Å, ±2.1°	±0.65 D	0.053	0.126
	(9s,6p)	[3s,2p]	(Na–Ar)	SCF: ±0.076 Å, ±1.1°				
	(12s,9p,3d)	[4s,3p,1d]	(K–Kr)	SCF: ±0.047 Å, ±1.1°				
	(15s,12p,6d)	[5s,4p,2d]	(Rb–Xe)	SCF: ±0.089 Å, ±1.2°[c]				
3-21G	(3s)	[2s]	(H–He)					
	(6s,3p)	[3s,2p]	(Li–Ne)	SCF: ±0.015 Å, ±4.8°	±0.017 Å, ±1.4°	±0.49 D	0.011	0.060
	(9s,6p)	[4s,3p]	(Na–Ar)	SCF: ±0.016 Å, ±3.3°				
3-21G$^{(*)}$	(9s,6p,1d)	[4s,3p,1d]	(Na–Ar)	SCF: ±0.017 Å, ±2.1°				
6-31G*	(4s)	[2s]	(H–He)					
	(10s,4p,1d)	[3s,2p,1d]	(Li–Ne)	SCF: ±0.014 Å, ±2.4°	±0.022 Å, ±1.1°	±0.30 D	0.002	0.060
				MP2: ±0.011 Å, ±1.7°	±0.013 Å, ±0.8°			
				MP4: ±0.012 Å, ±1.7°				
	(16s,10p,1d)	[4s,3p,1d]	(Na–Ar)	SCF: ±0.014 Å, ±2.2°				
				MP2: ±0.009 Å, ±1.2°				
6-31G**	(4s,1p)	[2s,1p]	(H–He)					
	(10s,4p,1d)	[3s,2p,1d]	(Li–Ne)	SCF: ±0.014 Å, ±2.4°			0.002	0.061
6-311G**	(4s,1p)	[3s,1p]	(H–He)					
	(10s,4p,1d)	[4s,3p,1d]	(Li–Ne)	SCF: ±0.012 Å, ±1.8°	±0.025 Å, ±1.3°			
				MP4: ±0.006 Å, ±1.9°	(±0.009)[d]			

DH SV[e]	(4s)	[2s]	(H–He)			
	(9s,5p)	[3s,2p]	(Li–Ne)	SCF: ±0.015 Å, ±5.7°	0.010	0.055
D DZP[f]	(4s,1p)	[2s,1p]	(H–He)			
	(9s,5p,1d)	[4s,2p,1d]	(Li–Ne)	SCF: ±0.009 Å, ±2.2°	0.011	0.060
MIDI[g]	(3s)	[2s]	(H–He)			
	(6s,3p)	[3s,2p]	(Li–Ne)	SCF: ±0.015 Å, ±3.2°	0.018	0.057

[a] AH_n errors for first-row compounds are based on 13 comparisons: H_2, LiH, CH_4, NH_3, H_2O, HF, BH, BH_2, BeH, CH_2, OH, CH, NH_2. The worst case for STO-3G is LiH, which is 0.086 Å longer than experiment. Ignoring LiH, the STO-3G average deviation falls to ± 0.028 Å. AH_n errors for second-row compounds are based on 5 comparisons: NaH, SiH_4, PH_3, H_2S, HCl. $A_lB_mH_n$ errors for first-row compounds are based on 24 comparisons: Li_2, LiOH, LiF, LiCl, B_2H_6, C_2H_2, C_2H_4, C_2H_6, HCN, HNC, CH_2NH, CH_3NH_2, CO, CO_2, H_2CO, CH_3OH, CH_3F, O_2, N_2, HNO, N_2H_2, N_2H_4, NH_2OH, HOF. Hehre et al. note that the error in A–B single bond lengths for a larger set of 30 molecules determined at the SCF level decreases in the order ±0.082 Å (STO-3G), ±0.067 Å (3-21G), ±0.040 Å (3-21G$^{(*)}$) and ±0.030 Å (6-31G*). The trend for multiple bonds was ±0.027 Å (STO-3G), ±0.017 Å (3-21G), ± 0.018 Å (3-21G$^{(*)}$), and ±0.023 Å (6-31G*). $A_lB_mH_n$ errors for third-row, main group compounds are based on 13 comparisons: KCN, KOH, GeH_3F, GeH_3Cl, Ge_2H_6, AsF, CH_3SeH, SeCO, LiBr, CH_3Br, Br_2, BrF, SiH_3Br. Dipole moment errors were based on 20 comparisons at the experimental geometries: CO, HCP, PH_3, ClF, H_2S, HCl, SiH_3F, NH_3, HF, H_2O, CH_3F, CH_3Cl, CS, H_2CO, HCN, LiH, NaH, LiCl, NaF, NaCl. The basis set superposition error (BSSE) is for H_2CO. The vertical ionization potential (VIP) error is the average error for ionizing to the $^2B_1$ and $^2B_2$ states of H_2CO. The VIP is the energy to remove an electron while keeping the molecular geometry fixed.

[b] Basis sets STO-3G through 6-311G** are Pople-type basis sets and are referenced in the text.

[c] Errors for third-row transition metal compounds are based on 9 comparisons: ScF_3, $TiCl_4$, VF_5, $VOCl_3$, CuF, CuCl, $Cr(CO)_6$, $Fe(CO)_5$, $Ni(CO)_4$.

[d] This estimate of the average error was obtained by scaling the MP2/6-31G* error for all 24 molecules by the ratio of the MP4/6-311G** error to the MP2/6-31G* error for the 6 molecules: Li_2, HCN, CO, CO_2, O_2, N_2.

[e] The Dunning–Hay "split valence" contracted basis.

[f] The Dunning "double zeta" contracted basis.

[g] The Huzinaga MIDI split valence contracted basis.

different results could have been obtained. For example, STO-3G underestimates the bond length in Na_2 by 0.72 Å and F_2 by 0.10 Å. The STO-3G basis is quite inflexible and gives a poor representation of both the atom and the molecule. Most of its apparent binding energy actually comes from a large BSSE.

With very small basis sets it is important to know which types of compounds and bonding situations are handled well and which are not. As basis sets become larger and more flexible their performance not only becomes better for "well-behaved" compounds, but also their ability to describe pathological cases improves even more. Thus, not only do split valence basis sets, such as 3-21G[44,45] and 3-21G$^{(*)}$[46] represent a marked improvement over STO-3G in terms of their ability to predict molecular structure (see Table 2), the errors for difficult to describe cases such as Na_2 and F_2 have dropped to 0.15 and 0.01 Å, respectively. 3-21G$^{(*)}$ is an extension of the 3-21G basis to second row elements (Na–Ar) and includes d polarization functions. In the case of 3-21G all this improvement in accuracy was achieved without increasing the number of Gaussian primitives relative to STO-3G, so the time required to compute the two-electron integrals has remained essentially the same.

Other split valence basis sets, such as the 4-31G, defined for first-[47] and second-row elements,[48] 5-31G, 6-31G,[49] and 4-21G[50] basis, differ only in the number of primitives used in expanding the inner shell and first contracted valence functions (2s and 2p). In particular, the 4-21G basis of Boggs and coworkers yields geometries that are often indistinguishable from the 3-21G results.

Because the error in *ab initio* bond lengths and angles tends to be systematic, it is possible to correct the computed geometrical parameters empirically. This procedure works best when applied to basis sets of at least split valence quality, since minimal basis set results are more erratic.

Attempts to improve the STO-3G wave function by introducing correlation recovery does *not* result in an overall improvement in the agreement with experiment. Since STO-3G errors are fairly evenly distributed, some positive and some negative, the introduction of correlation, which tends to uniformly lengthen bonds, may improve some but worsen others. For example, an STO-3G full CI improves the bond length in H_2 (from 0.71 to 0.74 Å where the experimental value is 0.74 Å), while the same treatment worsens the bond length in CO (from 1.15 to 1.19 Å where the experimental value is 1.13 Å).

With larger split valence polarization (SVP) basis sets, such as the 6-31G*, the introduction of correlation recovery via Møller–Plesset second- or fourth-order perturbation theory (MP2 or MP4) reduces the average error in bond lengths. Basis sets with even greater flexibility in the valence region can be constructed by allowing more Gaussian primitives to remain uncontracted and including polarization functions on both hydrogen and nonhydrogen atoms. One such example is the 6-311G**[51] basis, which was developed by optimizing Gaussian exponents and contraction coefficients at the MP2 level with polarization functions on both hydrogen and first-row atoms. As seen in Table 2, without correlation recovery the overall agreement with experiment is worse for $A_lB_mH_n$ compounds at the 6-31G* level than at the 3-21G level. As the Hartree–Fock limit is approached, most bond lengths are found to be too short. After incorporation of electron correlation effects, via perturbation theory or CI, most bond lengths tend to be closer to experiment, but somewhat too long. Therefore, efforts to try to improve split valence geometries by adding more basis functions, but keeping the simplicity of the HF wave function, are often misdirected.

Again, it is important to emphasize that in the preceding discussions we were dealing with *average* deviations from experiment. For any one particular molecule and any given sequence of wave functions the findings may be far different, as can be seen in the following example. The errors (ΔR_{CN}, ΔR_{CH} in Å) for the HCN molecule along the sequence (HF/STO-3G: 0.000, −0.005), (HF/3-21G: 0.016, 0.015), (HF/6-31G*: 0.020, 0.006), (MP2/6-31G*: −0.024, 0.005), (HF/6-311G**: 0.026, 0.007), (MP4/6-311G**: −0.031, 0.001), (MP3/6-311G**: 0.007, 0.001) start off low, rise steadily, and finally begin to fall only after quite large basis sets and sophisticated wave functions are employed. Since the last of these calculations consumes orders of magnitude more computer time than the initial STO-3G HF calculation, it might be concluded that, at least for HCN, one is better stopping with very simple theory. However, a slightly different sequence of wave functions (HF/4-31G, HF/6-31G**, MP3/6-31G**, MP3/6-311G**) listed in the book of Hehre et al. shows an almost monotonic improvement in ΔR with level of theory and would lead one to *exactly the opposite conclusion* concerning the effectiveness of more elaborate calculations. The conclusion to be drawn is that, on average, when larger basis sets are used in conjunction with correlation recovery methods (CI, perturbation theory, etc.), the overall agreement with experiment improves.

The data for other basis sets is far less complete. The Dunning–Hay[52] (9s,5p) → [3s,2p] split valence contraction provides about the same average error in AH_n geometry parameters as the Pople-style 3-21G basis. Similarly, Table 2 shows that the MIDI split valence basis of Huzinaga and co-workers[53] performs at about the same level as the 3-21G basis. The Dunning DZP basis,[54] which is slightly larger than the 6-31G* Pople basis, yields errors of the same order of magnitude. It would appear that, to the extent comparison data are available, the commonly used split valence or split valence + polarization basis sets all provide comparable levels of agreement with experimental geometry parameters.

Energy Differences

While an *ab initio* calculation can yield the total energy of a molecular system, chemists are ordinarily more interested in relative energies such as rotational barriers, conformational energy differences, or heats of reaction. It is, therefore, not necessary that a computed potential energy surface exactly match the "real" surface. All that is desired is a surface that would parallel the real surface to within some acceptable margin of error. The remarkable performance of the simple STO-3G basis in predicting molecular structure has been commented on previously. In some sense, the fact that the STO-3G restricted Hartree–Fock energy for a carbon atom is 406 kcal/mol higher than the exact energy is irrelevant. Some of this energy (~35 kcal/mol) is associated with the core electrons, which remain largely unaffected by changes in geometry along the potential surface. Because of this, semiempirical techniques and *ab initio* methods using effective core potentials can get away with ignoring the core completely. But the bulk of the STO-3G energy error is tied up with the (2s,2p) valence electrons. Perhaps the only advantage to be gained by employing methods capable of yielding a large fraction of the total energy is that there is less margin for error. The crucial question for people doing geometry optimizations is whether the STO-3G surface mimics the real surface sufficiently well in the vicinity of equilibrium structures that the resulting structural parameters are close to their experimental values.

Our discussion of the performance of various basis sets and methodologies in predicting energy differences will begin with the region of the potential energy surface near an equilibrium structure. Since we have already discussed the accuracy of optimal geometries, we can proceed to ask questions about the curvature of the surface at the minimum. The force constants, $\partial^2 V/\partial q_i \partial q_j$, are required for computing the vibrational modes of a molecule within the harmonic approximation. As in the case of molecular geometries, most of the calibration of basis sets with regard to normal mode analysis has been done by Pople and co-workers.

Unlike the situation with optimal geometries, the STO-3G minimal basis drastically overestimates vibrational frequencies in small molecules. The 30% deviation for CO, 2825 cm^{-1} compared to 2,170 cm^{-1} (experimental), is typical. Fortunately, the slightly larger 3-21G split valence set does much better, with most frequencies being overestimated by 10–15%.[55] Therefore, it is a common practice to apply an empirically determined correction factor of ~0.9 when comparing Hartree–Fock frequencies with experiment. Enlarging the basis set by introducing polarization functions, but remaining at the Hartree–Fock level, does not produce any improvement. However, when correlation recovery is introduced, as with MP2, the error drops almost in half,[56] even though computed frequencies are still larger than experiment.

The ability to obtain the complete set of vibrational modes for large polyatomic systems is of considerable importance. Experimentally this information is very difficult to determine and, once available, it becomes possible to compute thermodynamics quantities such as absolute entropies. Where necessary, improvements on the harmonic approximation have been computed by introducing cubic and quartic terms in studies of a variety of organic molecules.[57]

Looking more globally at the potential surface, we encounter phenomena such as rotational barriers and inversion barriers. The barriers to rotation in ethane and many other compounds are described well by any of the basis sets so far considered (STO-3G, 3-21G, 6-31G*), even at the HF level. However, when the barriers are small, as in the case of methanol (~1.1 kcal/mol), the HF method tends to overestimate ΔE_{rot}.

Inversion barriers, such as that found in ammonia, seem to require basis sets of at least 6-31G* quality or better. STO-3G barriers tend to be too large, often by a factor of two, whereas 3-21G barriers are too low. Once again, the performance of the minimal basis proves to be very irregular. If used in a carefully calibrated situation, it can produce results that are as good as calculations requiring orders of magnitude more computer time. But in the absence of demonstrated reliability, it cannot be trusted to provide any more than a qualitative indication of the answer.

Other global properties include considerations of energy differences across the entire potential energy surface. This includes bond breaking processes, heats of reaction, and the energies of transition states. *Ab initio* quantum chemistry has long sought to provide thermochemical data to an accuracy that is competitive with experiment. Although the accuracy of experimental data is sometimes hard to determine, it is approximately ±1 kcal/mol. Some authors[58] have recently suggested that this may be overly optimistic. Nonetheless, something on the order of ±1 kcal/mol is a reasonable goal for theoretical calculations, and we shall examine how well various basis sets approach it.

Table 3 shows the relative energies for four stationary points on the $CO_2 + H \rightarrow CO + OH$ potential energy surface. The reverse reaction is believed to be a major sink for carbon monoxide in the atmosphere. This reaction was chosen because the small size of the system allowed elaborate calculations to be performed. The only experimental piece of evidence regarding the reaction is the ΔH_r, which is 24 kcal/mol at 0°K. The experimental zero point vibrational contribution to ΔH is ~1 kcal/mol. Subtracting it will leave a target value of 23 kcal/mol for the energy differences between the molecules at their equilibrium geometries.

Hartree–Fock limit energies place the ΔE_r at 8 kcal/mol, but still endothermic. STO-3G makes a 41.7 kcal/mol error relative to the HF limit, finding the reactants *higher* in energy than the products. 3-21G does better at $\Delta E_r = -3.1$ kcal/mol and, finally, 6-31G** is within 2 kcal/mol of the HF limit. The agreement with experiment is typical of what is found in general. HF limit bond dissociation energies are often in error by 20–40 kcal/mol.

Second order Møller–Plesset perturbation theory, MP2, overestimates the effects of electron correlation by an amount that increases with the size of the basis (4 kcal/mol overestimate with STO-3G, 9 kcal/mol with 3-21G, and 11 kcal/mol with 6-31G**). Fourth-order perturbation, MP4, showed a much smaller error. The "exact" values are represented by the full CI entries under

Table 3 Energy Differences (in kcal/mol) at Four Stationary Points on the CO_2 + H → CO + OH Potential Surface[a]

	CO_2 + H	CO + OH	OCOH (SI)[b]	OC--OH (TS)[c]
STO-3G				
SCF	0.0	−34.1	−35.2	−14.4
MP2	0.0	−0.3	−2.4	11.5
MP4	0.0	−1.8		
Full CI	0.0	−4.3	−13.6	6.3
3-21G[d]				
SCF	0.0	−3.1	−1.1	13.9
MP2	0.0	29.9	13.7	31.9
Estimated full CI	0.0	20.9	−7.7	24.5
6-31G** [e]				
SCF	0.0	5.8	2.4	27.1
SD-CI	0.0	3.5		
Estimated SDQ-CI[f]	0.0	16.9		
MP2	0.0	31.8	4.4	34.3
MP4	0.0	24.0		
Estimated full CI	0.0	20.8 (25.2)[g]	14.3	28.9
Estimated HF limit[h]				
SCF	0.0	7.6		

[a] The difference is defined as E(arbitrary point) − $E(CO_2 + H)$. Thus, a positive value means that CO_2 + H is lower in energy. Experimentally, the zero point vibrationally corrected ΔE for the reaction is 23 kcal/mol. Geometries were taken from 6-31G** MCSCF optimized structures.

[b] Stable intermediate (trans form).

[c] Transition state (trans form).

[d] Full CIs could not be carried out for all four geometries with the 3-21G basis. Therefore, very large multireference CIs were performed.

[e] As with the 3-21G basis, full CIs could not be performed with the 6-31G** basis. Large multireference CIs, containing up to 345 configurations in the reference space and 398,533 configurations in the variational part of the calculation, were performed.

[f] An estimate for the contribution of quadruple excitations to the energy was made using the formula $\Delta E_Q = \Delta E_{SD}(1 - c_0^2)$ where ΔE_{SD} is the singles and doubles energy lowering and c_0 is the coefficient of the Hartree–Fock configuration in the CI wave function.

[g] Energy difference based on the supermolecule approach. For the STO-3G and 3-21G basis sets, a full CI could actually be performed on CO and OH fragments. With the 6-31G** basis a full CI was too large to do. The 4.4 kcal/mol difference is a measure of the size consistency uncertainty in the estimated MR SDQ-CIs on which the 6-31G** entries are based.

[h] The estimated complete basis set limit is based on numerical Hartree–Fock calculations for CO and OH and a very extended basis calculation on CO_2 at 2.192 bohr, which gave an SCF energy of −187.7244 hartree.

each basis. Actual full CI calculations could be performed only for the STO-3G basis and the CO and OH molecules in the 3-21G basis. The remaining "Full CI" entries were estimated from large multireference single and double excitation CIs (MR SD-CIs). The effects of quadruple excitations were estimated through the use of a multireference analog of the so-called "Davidson" or "size-consistency" correction [labeled MR SD(Q)-CI],[59] but the magnitude of the correction was small. The largest variational CI involved nearly 400,000 configurations and had a reference space of 345 configurations.

An interesting sidelight of the results shown in Table 3 has to do with the relative merits of configuration interaction compared to perturbation theory as a method for computing energy differences across a potential surface. One desirable characteristic of Møller–Plesset perturbation theory is that it is "size consistent," i.e., the energy of *n* infinitely separated systems is just *n* times the energy of a single system. Although the HF wave function is size consistent, the single- and double-excitation CI (SD-CI) wave function is not. Moreover, as the size of system grows the magnitude of the size consistency error associated with SD-CI also grows.

In fact, the SD-CI ΔE_r with the 6-31G** basis is even smaller than the HF result (a 20 kcal/mol error), while the MP2 result is only 11 kcal/mol in error (but in the other direction). However, this comparison is somewhat artificial, since few people would stop with the raw SD-CI energy difference. At minimum, the "size consistency" correction for the effect of quadruple excitations is usually added to the SD-CI answer. This entry is labeled "est. SD(Q)-CI" in Table 3 and is seen to be only 4 kcal/mol less than the full CI limit. Other full CI calculations[60] have demonstrated that this simple correction works surprisingly well for small molecules. If less reliance on this correction is desired, a whole host of MR SD-CIs can be employed that gradually approach the full CI result. However, the size consistency error in the estimated full CI result for CO + OH computed as a supermolecule compared to the sum of the energies of CO and OH computed separately was still ~4 kcal/mol for the largest MR SD(Q)-CI wave function.

Size consistency is only one of a number of desirable characteristics in a theoretical model. In and of itself, it does not guarantee accuracy in the total energy or any other property. A method that is not strictly size consistent may, in fact, more closely match the full CI value for certain properties in certain molecules. The user must be wary of the potential uncertainty associated with *any* theoretical calculation.

Further illustrations of the effects of basis set enlargement and correlation recovery are given in Table 3 for a stable intermediate geometry and a transition state geometry. The errors are seen to be roughly comparable to the errors associated with the energy difference between reactants and products.

Even with polarized basis sets, such as 6-31G**, and fourth-order perturbation theory, the errors in bond dissociation energies can be unacceptably large for some purposes. Extension of the basis through the addition of diffuse s and p functions, a second set of d-type polarization functions, and introduction of f functions on Li–Ne and d functions on hydrogen leads to a basis set with the almost unpronounceable name of 6-311++G(3df,3pd).[61] At the MP4 level this basis set produces superior dissociation energies to 6-31G**. For example, the error in the dissociation energy of H_2 drops from 3.9 to 1.3 kcal/mol. For the dissociation of N_2, the error drops from 23.7 to 4.8 kcal/mol, and for $N_2 + 3H_2 \rightarrow 2NH_3$ the error decreases from 15.1 to 1.2 kcal/mol. Pople and co-workers have suggested that computed energy differences can be improved by assuming the additivity of basis set enlargement effects and correlation effects. They propose avoiding the expense of large basis set/high correlation recovery calculations by performing several smaller calculations. Although additivity cannot be justified on purely formalistic grounds, these authors note that experience has shown this approach to work.

One-Electron Properties

There has been no systematic compilation of data on the performance of most popular basis sets for one-electron properties other than the dipole moment. Table 2 provides a rough indication of how well the Pople-style basis sets do at the HF level. The effect of electron correlation on the dipole moment (and other properties) varies greatly from molecule to molecule. An indication of the effect of increasing basis set size can be seen in the following data for the water molecule at the experimental geometry: HF/STO = 1.73 D, HF/3-21G = 2.44 D, HF/6-31G* = 2.22 D, HF limit = 1.98 D. Large-scale CI brings the value down to 1.87 D compared to 1.85 D (experimental). The review by Davidson and Feller[5] lists results for 11 different basis sets and a large number of properties on formaldehyde.

IN-DEPTH DISCUSSION

We now return to several of the topics already introduced, to cover them more in-depth and, then, proceed to examine the topic of weak intermolecular interactions.

Sources of Gaussian Primitives and Contraction Coefficients

Early use of Gaussians saw much experimentation with ways to choose the exponents. Sometimes they were optimized for a particular molecule, as had been done with exponential functions. However, the increased number of Gaussians, relative to STOs, made this impractical.

Initial efforts to derive transferable sets of exponents resulted in the first energy-optimized Gaussian basis sets for atoms, published by Huzinaga.[62] For each of the first-row elements, the SCF energy for a particular number of s- and p-type functions was optimized with respect to each of the Gaussian primitive exponents. Since the largest set included 10s and 6p functions, this meant a 16-dimensional space had to be searched for a minimum in E_{SCF}. Dunning produced contracted basis sets from both the (9s,5p) and (10s,6p) sets of exponents.[63]

Van Duijneveldt[64] extended Huzinaga's work through (14s,9p). This basis produced an error in the neon SCF energy of only 0.00023 hartree, compared to 0.0060 hartree for Huzinaga's (10s,6p) set. The difficulties associated with independently optimizing each exponent in large Gaussian basis sets are associated with the flatness of the surface and near linear dependence in the variational parameters. This is illustrated by the fact that Faegri[65] was able to obtain another 0.00008 hartree from the van Duijneveldt (14s,9p) neon basis by employing more refined optimization methods. Such difficulties were responsible for the lack of even larger independently optimized basis sets appearing in print until the very recent work of Partridge on second-row (Na–Ar)[66] and first-row (Li–Ne) plus third-row (K–Kr) atoms.[67] The most accurate of Partridge's basis sets (18s,13p) are within 4 μhartrees of the numerical Hartree–Fock energy for the first-row atoms. The corresponding errors for the second- and third-row atoms were 25 and 30 μhartrees, respectively, with basis sets of (20s,15p) and (24s,16p,12d).

Even-Tempered Gaussians

Shortly after van Duijneveldt published his list of energy-optimized Gaussian exponents, a careful inspection of these values led Ruedenberg and coworkers[68] to revive an earlier suggestion of Reeves[69] that the ratios of successive exponents be held constant. Ruedenberg attached the label "even-tempered" to such sets and published a series of papers on both exponential and Gaussian type basis sets.[70–72] In an even-tempered basis the ith exponent, ζ_i, is given by $\alpha\beta^i$, where α and β are fixed numbers for each s,p,d . . . symmetry. If independently optimized exponents are plotted on a log scale the spacing grows slightly wider at both the diffuse end (small ζ) and at the tight end (large ζ), but remains nearly constant throughout most of the exponents. An even-tempered set forms a geometric sequence with all exponents being equidistant from their neighbors on a $\log(\zeta)$ plot.

The chief initial advantage to be gained from the even-tempered approximation is the reduction in the number of parameters to be optimized. For example, all exponents in an (s,p) basis set on carbon can be specified in terms of only four parameters (α_s, β_s, α_p, and β_p) *regardless of the number of primitives.*

It was subsequently discovered that the optimized $\alpha(N)$ and $\beta(N)$, where N is the number of primitives, can be parameterized as functions of N. This provided a way of extrapolating to obtain values of α and β appropriate for very large basis sets without the need for an expensive reoptimization.[73] Schmidt and Ruedenberg[74] published lists of α and β for elements up through Ar so that, with the addition of polarization functions, it is now possible to easily generate a sequence of wave functions that smoothly approaches a complete set.

An energy penalty must, of course, be paid for imposing the even-tempered restriction. As an example, an independently optimized (8s,4p) basis gives an energy for the carbon atom of −37.680 hartree, while an even-tempered (8s,4p) set gives only −37.668. If one additional s function is added to the even-tempered set, the energy falls to −37.677 hartree, so the penalty was worth about one s function. At the large basis set end of the spectrum, the cumulative effect of all the adjustable parameters in an independently optimized set seems to be worth many more even-tempered primitives. Partridge lists an energy of −37.688619 for carbon with an (18s,13p) basis, whereas it takes a (26s,13p) even-tempered basis to achieve the same energy.

For correlation purposes the basis set which gives the lowest SCF energy may not give the lowest CI energy. For example, an independently optimized (19s,14p) basis set, which included an extra (s,p) diffuse set, gave a slightly lower (0.00001 hartree) oxygen atom SCF energy than a (23s,12p) even-tempered set. The two basis sets contain approximately the same total number of functions (61 for the independently optimized set compared to 59 for the even-tempered set). But at the SD-CI level, the even-tempered set yielded slightly more correlation energy (by 0.00002 hartrees).

The ease of generating arbitrarily large even-tempered sets has meant that some of the lowest energy SCF and CI wave functions for first-row atoms and many small molecules have been produced with these basis sets. Tables 4 and 5 illustrate the sorts of results that may be achieved.

Wilson and co-workers have proposed using a large enough even-tempered primitive set that the same set of exponents can be used for all first row elements.[75,76] Of course, this requires very large basis sets if neon is to be treated with the same accuracy as lithium.

The concepts of even-tempered exponents and universal basis sets were combined along with six constraints to produce "geometrical basis sets"[77] for all atoms from H to Sr. The same set of exponents (but differing numbers of s, p, and d functions) appears. Errors in the atomic energies range from 0.0002 hartree for the 6s hydrogen basis to 0.0005 hartree for the (13s,8p) carbon basis to 0.0555 hartree for the (20s,13p,8d) Sr basis.

Well-Tempered Gaussians

Huzinaga and co-workers[78] have extended the even-tempered idea by developing a four-parameter function of the form:

$$\zeta_i = \alpha\beta^{k-1}[1 + \gamma(k/K)^{\delta}], \qquad k = 1, 2, \ldots, K$$

Table 4 Examples of Extended Even-Tempered Gaussian Basis Set CI Calculations on Atoms[a]

	B (^{2}P)	C (^{3}P)	N (^{4}S)	N$^+$ (^{3}P)	O (^{3}P)	O$^-$(^{2}P)	F (^{2}P)
Date	1987	1987	1987	1987	1988	1988	1987
E_{SCF}	−24.5291	−37.6886	−54.4009	−53.8880	−74.8094	−74.7897	−99.4093
E_{CI}	−24.6511	−37.8410	−54.5835	−54.0499	−75.0569	−75.1052	−99.7163
#Selected	39,852	84,611	157,632	75,443	209,036	305,016	136,129
#Total	446,837	999,574	2,001,192	1,124,191	5,275,566	13,313,934	3,952,228
%E_{corr}	97.6	97.5	96.7	96.3	95.4	94.8	94.3
E_{val}	−24.6016	−37.7884	−54.5277	−53.9949	−74.9984	−75.0477	−99.6555
Ref. space	32	36	42	41	81	225	71
$\sum c^2$	0.98	0.97	0.99	0.99	0.99	0.98	0.96
$\|C_{max}\|$	0.012	0.014	0.017	0.010	0.012	0.012	0.014
A_{iso}	6.4	17.8	10.1	−30.2	285.3		
A_{iso} (exp.)	11.6	10.4	−34.5		301.7		

[a] All atomic calculations, except oxygen, were done with a (23s,12p,10d,4f,2g) uncontracted even-tempered basis, for a total of 155 functions. For O and O$^-$ one more set of diffuse f and g functions were added, for a total of 171 functions.

Date	Year in which the calculation was performed.
E_{SCF}	The self-consistent field energy in hartrees.
E_{CI}	Total CI energy (no cores) in hartrees.
#Selected	Number of spin-adapted CI configurations selected by perturbation theory.
#Total	Total number of spin-adapted configurations before selection.
%E_{corr}	Percentage of the empirical correlation variationally recovered.
E_{val}	Valence CI energy (1s electrons are core).
Ref. space	Number of configurations in the CI reference space.
$\sum c^2$	Sum of the squares of the reference configurations in the CI wave function. This is a rough measure of the quality of the reference space.
$\|C_{max}\|$	The absolute value of the largest CI coefficient appearing outside the reference space. For multiple spin couplings $\|C_{max}\|$ is taken as a sum over all couplings associated with a given space orbital product, $\|C_{max}\|^2 = \sum c_i^2$. $\|C_{max}\|$ provides a measure of the quality of the reference space.
A_{iso}	The isotropic hyperfine property, $(8\pi/3)(g_e/g_o)g_N\beta_N\langle\delta(O)\rangle_{spin}$, in MHz. The nuclear g values used in converting from atomic units to MHz were as follows: B 1.7923, C 1.4048, N 0.4038, O −0.7575, F 5.2576. The gas phase experimental values are taken from references cited in the original articles.

Table 5 Examples of Extended Even-Tempered Gaussian Basis Set CI Calculations on Small Molecules[a,b]

	HF (117 funct.)	OH (117 funct.)	H_2O (140 funct.)	CO (136 funct.)
Date	1988	1989	1987	1987
E_{SCF}	−100.0706 (−100.0707)	−75.4212 (−75.4213)	−76.0673	−112.7903 (−112.7909)
E_{CI}	−100.4251 (−100.460)	−75.7084 (−75.732)	−76.3963 (−76.400)	−113.2193 (−113.326)
#Kept	94,773	106,321	33,788	88,618
#Total	2,405,746	4,125,343	3,899,802	7,996,821
$\%E_{corr}$	91	92	88	80
E_{val}	−100.3661	−75.6514	−76.3427	
Ref. space	30	68	15	51
$\sum c^2$	0.98	0.98	0.96	0.96
$\langle \mu_z \rangle$	−0.714 (−0.707)	0.645 (0.656)	0.736 (0.727)	−0.040 (−0.048)
Θ	1.704 (1.7)	1.230 (1.217)	1.900 (1.960)	−1.522
$\langle r^2 \rangle_e$	13.969	15.44	19.66 (18 ± 2)	39.79
$\langle q_z/r^3{}_A \rangle$	0.0003 (0.0)	−0.0001 (0.0)	0.0000 (0.0)	0.0024 (0.0)
$\langle q_z/r^3{}_B \rangle$	0.0035 (0.0)	−0.0027 (0.0)	−0.0001 (0.0)	−0.0013 (0.0)
A_{iso} (O)		−39.5 (−44.9)		
A_{iso} (H)		−68.3 (−73.1)		

	CO_2 (201 funct.)	H_2COH^+ (126 funct.)	NH_3 (130 funct.)
Date	1989	1988	1987
E_{SCF}	−187.7244	−113.5670	−56.2246
E_{CI}	−188.4108	−75.7084	−56.4946 (−56.563)
#Kept	264,143	143,970	19,936
#Total	32,874,540	24,546,552	1,461,268
Ref. space	45	142	10
$\sum c^2$	0.92	0.95	0.91
$\langle \mu_z \rangle$		−1.287	−0.613 (−0.579)
Θ	−3.315	1.230	2.183 (2.42)
$\langle r^2 \rangle_e$	113.107		26.71 (25.50)
$\langle q_z/r^3{}_A \rangle$	0.0029 (0.0)		0.0012 (0.0)
$\langle q_z/r^3{}_B \rangle$			−0.0009 (0.0)
A_{iso} (H)		253 (372)	
A_{iso} (C)		−79 (−109)	

[a] The entries E_{SCF}, E_{CI}, #Selected, #Total, %E_{corr}, E_{val}, Ref. space, and $\sum c^2$ have the same definitions as in Table 4. Values in parentheses are either the experimental value or, in the case of E_{SCF}, the numerical HF value. Properties were obtained from the CI wave function.

[b] The HF basis was (23s, 12p, 6d, 2f/10s, 4p, 2d) → [14s, 10p, 6d, 2f/7s, 4p, 2d]. R_{HF} = 0.917 Å. The OH basis was (23s, 12p, 6d, 2f/10s, 4p, 2d) → [14s, 10p, 6d, 2f/7s, 4p, 2d]. R_{OH} = 0.971 Å. The H_2O basis was (27s, 16p, 7d, 2f/19s, 7p, 3d) → [12s, 10p, 4d, 2f/10s, 4p, 2d]. R_{OH} = 0.957 Å, 104.5°. The CO basis was (19s, 10p, 4d, 2f) → [10s, 6p, 4d, 2f]. R_{CO} = 1.128 Å. The CO_2 basis was (19s, 10p, 5d, 2f) → [10s, 6p, 5d, 2f]. R_{CO} = 1.160 Å. The H_2CO^+ basis was (19s, 10p, 3d, 1f/10s, 2p, 1d) → [10s, 5p, 3d, 1f/5s, 2p, 1d]. R_{CO} = 1.198 Å, R_{CH} = 1.117 Å, 123.0°. The origin was placed at the center of mass. The NH_3 basis was (19s, 10p, 3d, 1f/10s, 3p, 2d) → [10s, 8p, 3d, 1f/6s, 3p, 1d]. R_{NH} = 1.012 Å, 106.7°.

for generating "well-tempered" Gaussian exponents. Since the exponents are shared over s,p,d,f . . . Gaussians the number of independent parameters does not increase as higher angular momentum functions are included. Optimizing α, β, γ, and δ produced a (14s,9p) SCF energy for the neon atom only 0.1 mhartree higher than the independently optimized set. As with the even-tempered primitives, these exponents are energy optimized in atomic SCF calculations.

Besides a (14s,9p) set for the first row elements, a (16s,11p) set was reported for second-row elements. Subsequent papers reported basis sets for the third-row[79] and second-row transition metals.[80] SCF excitation energies for various s^2d^{n-2}, s^1d^{n-1}, and d^n states were within 0.04 eV (first-row transition elements) and 0.01 eV (second-row transition elements) of the numerical Hartree–Fock values.

MINI-*i*, MIDI-*i*, and MAXI-*i*, etc.

Among other rich sources of Gaussian exponents and contraction coefficients are the books by Poirier et al.[81] and Huzinaga and co-workers.[53] Huzinaga and Sakai[82] reported Gaussian exponents and contraction coefficients for elements Li–Ar as long ago as 1969. This work was continued by Tatewaki and Huzinaga[83] in a large collection of minimal and split valence basis sets for all first row atoms. Contraction coefficients and exponents were optimized at the SCF level. The MINI-*i* (i = 1–4) sets all possess three Gaussians in the 2s function and have varying numbers of Gaussians in the 1s and 2p contracted functions. Using the notation (3s,3s/3p), MINI-1 is (3,3/3), MINI-2 is (3,3/4), MINI-3 is (4,3/3), and MINI-4 is (4,3/4). By splitting the three-term valence space s and p contractions into one function with two primitives and the other primitive by itself, you obtain the MIDI-*i* sets, where, for example, MIDI-1 is a [3s,2s,1s/2p,1p] contraction.

The common denominator of the MINI and MIDI contractions is the use of three terms to represent the 2s atomic orbital. In the MAXI-*i* sets[84] four Gaussian primitives are used along with up to seven Gaussians for the 1s and 2p functions. These may be split in a variety of ways to yield basis sets, such as [7s,2s,1s,1s/4p,1p,1p,1p], which can be used in conjunction with polarization functions to yield high-accuracy results. No extensive calibration has been done for these basis sets.

The first article in the series of papers by Tatewaki and Huzinaga dealt with first-row transition metals (Sc–Zn).[85] They reported two new minimal basis sets, originally called STD–SET(1) and DZC–SET(1), which yielded atomic orbital energies as good as, or better than, available double zeta quality basis sets. The same authors[86] suggest procedures for splitting the valence shell AO so as to increase the basis set's flexibility.

Other articles in this series discuss the effect of the s-symmetry combination of Cartesian d's on molecular binding energies,[87] and other contracted basis sets for elements through the second-row transition metals.[88,89]

Tests of the STO-3G basis set on transition metal compounds demonstrated that, at least as far as the orbital energies of the 3d electrons and molecular geometries are concerned, this minimal basis performed rather poorly. It had also been necessary to introduce 4p (for first-row) and 5p (for second-row) functions to obtain qualitatively good results in molecules. In the same (1s,2s,3s,4s/2p,3p/3d) notation introduced previously for the MINI-*i* basis sets, the MINI-1 sets for Sc–Zn are denoted as (3,3,3,3/3,3/3). Not only are the MINI-1 sets smaller than the corresponding STO-3G sets, but the resulting 3d orbital energies are nearly as good as a double-zeta quality basis set.[90] A comparison of the MINI-1's performance in geometry optimizations on nine transition metal compounds reveals it to be as good as, or better than, the STO-3G sets. The worst performance, relative to STO-3G, occurs in C—O bond lengths, which are ~0.07 Å too long. This problem is solved if split valence quality contractions are used for CO.

Still Others

McLean and Chandler[91] have developed a widely used set of minimal, double- and triple-zeta contracted functions for second-row elements Na–Ar from Huzinaga's (12s,8p) and (12s,9p) primitive sets.

Atomic Natural Orbitals

In the previous discussion of contracted basis functions, the "general" contraction scheme of Raffenetti, in which every primitive can contribute to every contracted function, was mentioned. The alternate approach produces functions that are sometimes referred to as "segmented" contractions since the primitives are partitioned into disjoint or almost disjoint sets. Because FORTRAN programs to efficiently implement the general contraction scheme were not very common in the 1960s and 1970s, most basis sets are of the segmented variety.[92] For example, all of the Pople-style basis sets are of this type.

A recent suggestion by Almlöf and Taylor[93] takes advantage of the general contraction scheme to construct contracted basis functions from the atomic natural orbitals (ANOs) obtained from valence SD-CI calculations on the free atoms. They argue that large primitive sets can be contracted in this manner without significant loss of accuracy in either SCF or correlated calculations. The natural orbitals are obtained as the eigenvectors of the first-order density matrix and their associated eigenvalues are referred to as the orbital "occupation numbers." Such numbers provide a measure of the relative importance of each orbital in expanding the density matrix.

Almlöf and Taylor propose using all ANOs with comparable occupation numbers. Thus, in choosing a contracted basis set one would first decide approximately how many functions to include. Second, one would examine a chart of the occupation numbers for the particular atoms involved in the molecule and select some threshold such that all ANOs with larger occupations would be included. For example, when ANO occupation numbers are used as a criterion for how to expand the minimal basis in neon the following sequence is obtained: [2s,1p], [3s,2p,1d], [4s,3p,2d,1f], [5s,4p,3d,2f,1g]. The authors conclude that the last of these sets, the (13s,8p,6d,4f,2g) → [5s,4p,3d,2f,1g] basis, suffers no more than 1 kcal/mol contraction error for each angular momentum type. A study by Ahlrichs and co-workers[94] of the impact of higher polarization functions on F_2 found roughly the same relative importance of various types of polarization functions as was found by Almlöf and Taylor.

In subsequent papers by Almlöf et al.,[95,96] the ANO basis sets were applied to the calculation of molecular properties by supplementing the original energy-optimized sets with diffuse functions, in accord with the experience of others[97] that such functions are often required for good agreement with experiment. Additionally, some of the functions were permitted to float independent of the atomic centers. Their positions were determined by minimizing the SCF energy using a second-order convergence scheme and analytical first derivatives. The motivation for freeing the functions from the atomic centers was somewhat different from the motivation behind the subminimal floating spherical Gaussians basis sets used in the 1960s. In this instance, the intent is to better approximate the shift in the center of gravity of the charge distribution resulting from bond formation.

These authors maintain that properties, such as the dipole moment, polarizabilities, and infrared intensities, can be reproduced with near Hartree–Fock accuracy using a relatively small DZP + diffuse basis. In the review by Davidson and Feller[5] the results from an atom-fixed ANO basis are compared with results obtained from many other types of basis sets for a variety of energy-related and 1-electron properties of formaldehyde.

Other recent applications of ANO basis sets include the work of Bauschlicher and Langhoff on the $^4\Delta$ ground state of FeH.[98] Large multiconfiguration SCF (MCSCF)/MR SD-CI calculations were carried out with a (20s,15p,10d, 6f,4g/8s,6p,4d) primitive basis contracted to [8s,7p,5d,3f,2g/4s,3p,2d]. Exponents for the f and g functions were chosen in an even-tempered fashion ($\alpha_f = 0.0696$, $\alpha_g = 0.2088$, $\beta_{f,g} = 2.5$).

The principal disadvantage of the above-mentioned ANO approach to generating basis functions is the large number of primitives required, since the integral evaluation time goes like n^4, where n is the number of Gaussian primitives. Based on a study of the oxygen atom, Dunning[99] recently proposed smaller sets of primitives in a "correlation consistent" series of basis sets that yield 99% of the correlation energy of the ANO sets while using 33% fewer primitives. Like previous authors[100] Dunning chose to optimize (d,f,g) exponents at the SD-CI level using an even-tempered restriction. For the (s,p) portion of the basis, he demonstrated that the HF (1s,2s,2p) set, when augmented with diffuse primitives as suggested by Raffenetti, generally does as well as the ANO (s,p) basis. Both the ANO and correlation consistent basis sets were developed to treat the valence portion of the correlation energy.

Several other authors have also pointed out advantages to floating orbitals. Huber's floating orbital geometry optimization (FOGO) method[101] applies the Hellmann–Feynman force to the nuclei and the energy gradient to the basis function centers to obtain a wave function that has a vanishing gradient and a vanishing Hellmann–Feynman force.

Some applications of perturbation theory to molecular problems would benefit from the simplicity of an extended floating spherical Gaussian basis. Adamowicz and Bartlett[102] have developed a procedure for projecting large conventional basis set wave functions onto a floating spherical Gaussian basis.

Miscellaneous basis sets for first- and second-row elements come from Roos and Siegbahn[103] and for first-row transition metal atoms (Sc–Zn) from Faegri and Speis.[104] The latter include energy optimized sets ranging from (12s,6p,4d) up to (16s,11p,8d). A (15s,8p,5d) basis for first-row transition metals was developed by Basch et al.[105] Still other sets are due to Wachters[106] and Roos et al.[107] Hay[108] argued that the flexibility of Wachters' d-space set should be increased. He has shown that the 3d orbitals of the $4s^23d^{n-2}$, $4s^13d^{n-1}$, and $3d^n$ electronic configurations are sufficiently different that the use of basis sets optimized for one can lead to errors of several electron volts in computed excitation energies relative to the Hartree–Fock limit. Comparable problems do not arise for first-row atoms, since the 2s and 2p orbitals for $2s^22p^n$ and $2s^12p^{n+1}$ configurations are very similar to each other.

Functions for Augmenting Basis Sets

The role of polarization functions in expanding the flexibility of split valence or double-zeta quality basis sets has already been discussed. One of two approaches is normally followed in determining reasonable values for the exponents of such functions. Since polarization functions do not contribute to the HF occupied orbitals of the atom's ground state, their exponents cannot be determined in the same straightforward manner as, for example, the (s,p) exponents for carbon. In the 1960s it was common to optimize polarization function exponents for a particular molecule (often a di- or triatomic) under examination. As *ab initio* calculations on larger molecules were attempted, the amount of time needed to reoptimize polarization exponents for each new molecule was judged to be excessive. A search for some common set was begun.

With the 6-31G** basis sets, Pople and co-workers recommended a set of polarization function exponents based on an examination of the optimal SCF values for a variety of molecules. In most instances, the energy loss that resulted from not using the optimal polarization exponent was small because the total energy was a slowly varying function of ζ_{pol}. Other authors have taken the same approach.

Although polarization function exponents do not contribute to HF AOs, they will make a contribution to the atomic wave function if correlation recovery is introduced. For example, d functions account for over 40% of the correlation energy in nitrogen. Thus, as already mentioned in regard to Dunning's "correlation consistent" basis sets, some researchers have optimized polarization exponents at some low level of atomic CI or perturbation theory and then used these values in molecular calculations. Representative values of polarization exponents for some elements are listed in Table 6. A feeling for the contribution of these functions to the total energy can be obtained from Table 7. It is also apparent that the contributions from higher L functions become more important when the relative importance of radial correlation/angular correlation shifts in favor of the latter as Z increases.

An alternative to introducing higher L functions into the basis is to add lower angular momentum functions at the bond centers, where the build-up of charge is greatest.[109] This approach borrows something from the floating basis function idea, but does not allow the functions to be placed arbitrarily in space. For first-row diatomics the introduction of a single (s,p) set at the center of the bond provides 90% of the energy lowering obtained with a single set of d functions at the nuclear centers. Many authors have suggested exponents for these functions and have advocated their use in SCF and correlated wave function calculations.[110–112]

When conventional atom-centered basis sets are used to compute the SCF dissociation energy of a first-row diatomic molecule, D_e, is usually underestimated. This is because the HF limit for the separated atoms is relatively easy to approach, requiring only s and p functions, whereas the molecule will require many additional higher L functions at R_e. The tendency to favor the separated atoms persists even at the CI level. The introduction of bond-centered functions, which contribute to the molecule but not the separated atoms, has been proposed by Wright and Williams.[113,114] By exploiting the basis set superposition error at R_e and balancing it against the basis set incompleteness error at $R = \infty$, the authors report excellent agreement with experiment for molecules like HF. Wright and Buenker[115] have argued that systematically improving both the atom-centered and bond-centered basis sets should allow D_e to be better estimated by simultaneously converging from above and below. On the other hand, Bauschlicher[116] has warned that this balance is very difficult to maintain as the geometry and correlation recovery method are varied. Further evidence of the kind of accuracy which can be expected from bond functions comes from an analysis of 14 single- and multiple-bonded molecules at the MP2 and MP3 levels. The 6-31G* basis is compared with the 6-31G basis with added bond functions (bf). The observed root mean square deviations, in kcal/mol, are MP2: 8.1 (bf), 13.4 (6-31G*); MP3 10.3 (bf), 20.0 (6-31G*).[117]

Table 6 Selected Polarization Function Exponents

		Source
Hydrogen		
1p:	1.10	6-31G**, average of optimal SCF values[a]
	0.97	Optimal SCF values for H_2 (Feller)[b]
	0.727	Optimal full CI value for H_2 (Dunning)[c]
2p:	(0.38, 1.62)	Optimal SCF values for H_2
	(0.388, 1.407)	Optimal full CI values for H_2 (Dunning)
3p:	(0.292, 0.838, 2.292)	Optimal full CI values for H_2
1d:	1.47	Optimal SCF value for H_2
	1.057	Optimal full CI value for H_2 (Dunning)
	2.00	Pople and co-workers recommended value[d]
2d:	(0.662, 2.062)	Optimal full CI values for H_2 (Dunning)
1f:	1.397	
Helium		
1d:	2.00	Pople and co-workers recommended value
Lithium		
1d:	0.20	6-31G**, average of optimal SCF values
1f:	0.15	Pople and co-workers recommended value
Beryllium		
1d:	0.40	6-31G**, average of optimal SCF values
1f:	0.26	Pople and co-workers recommended value
Boron		
1d:	0.60	6-31G**, average of optimal SCF values
	0.343	Optimal valence SD-CI value for the atom
2d:	(0.199, 0.661)	
3d:	(0.145, 0.402, 1.110)	
1f:	0.490	
2f:	(0.311, 0.882)	
1g:	0.673	
1f:	0.50	Pople and co-workers recommended value
Carbon		
1d:	0.80	6-31G**, average of optimal SCF values
	0.550	Optimal valence SD-CI value for the atom
2d:	(0.318, 1.097)	
3d:	(0.228, 0.649, 1.848)	
1f:	0.761	
2f:	(0.485, 1.419)	
1g:	1.011	
1f:	0.80	Pople and co-workers recommended value
Nitrogen		
1d:	0.80	6-31G**, average of optimal SCF values
	0.817	Optimal valence SD-CI value for the atom
2d:	(0.469, 1.654)	

(*continued*)

Table 6 Selected Polarization Function Exponents (*continued*)

	Source
Nitrogen (*continued*)	
3d: (0.335, 0.968, 2.837)	
1f: 1.093	
2f: (0.685, 2.027)	
1g: 1.427	
1f: 1.00	Pople and co-workers recommended value
Oxygen	
1d: 0.80	6-31G**, average of optimal SCF values
1.185	Optimal valence SD-CI value for the atom
2d: (0.645, 2.314)	
3d: (0.444, 1.300, 3.775)	
1f: 1.428	
2f: (0.859, 2.666)	
1g: 1.846	
1f: 1.40	Pople and co-workers recommended value
Fluorine	
1d: 0.80	6-31G**, average of optimal SCF values
1.640	Optimal valence SD-CI value for the atom
2d: (0.855, 3.197)	
3d: (0.586, 1.725, 5.014)	
1f: 1.917	
2f: (1.148, 3.562)	
1g: 2.376	
1f: 1.85	Pople and co-workers recommended value
Neon	
1d: 0.80	6-31G**, average of optimal SCF values
2.202	Optimal valence SD-CI value for the atom
2d: (1.096, 4.014)	
3d: (0.747, 2.213, 6.471)	
1f: 2.544	
2f: (1.524, 4.657)	
1g: 2.983	
1f: 2.50	Pople and co-workers recommended value
Sodium	
1d: 0.175	6-31G**, average of optimal SCF values
1f: 0.15	Pople and co-workers recommended value
Magnesium	
1d: 0.175	6-31G**, average of optimal SCF values
1f: 0.20	Pople and co-workers recommended value
Aluminum	
1d: 0.325	6-31G**, average of optimal SCF values
0.19	Optimal valence SD-CI value (Feller)
2d: (0.24, 4.84)	Optimal SD-CI value, no core (Feller)
(0.11, 0.34)	Optimal valence SD-CI value (Feller)
1f: 0.25	Pople and co-workers recommended value

(*continued*)

Table 6 Selected Polarization Function Exponents (*continued*)

		Source
Silicon		
1d:	0.45	6-31G**, average of optimal SCF values
	0.28	Optimal valence SD-CI value (Feller)
2d:	(0.31, 5.90)	Optimal SD-CI value, no core (Feller)
	(0.15, 0.46)	Optimal valence SD-CI value (Feller)
5d:	(0.07, 0.20, 0.59, 1.86, 7.30)	Recommended SCF values (Magnusson and Schaefer)
1f:	0.32	Pople and co-workers recommended value
	0.35	Optimal valence SD-CI value (Feller)
	7.30	Optimal SD-CI value, no core (Feller)
Phosphorus		
1d:	0.55	6-31G**, average of optimal SCF values
	0.38	Optimal valence SD-CI value (Feller)
2d:	(0.47, 9.03)	Optimal SD-CI value, no core (Feller)
	(0.21, 0.63)	Optimal valence SD-CI value (Feller)
5d:	(0.08, 0.22, 0.63, 1.99, 7.85)	Recommended SCF values (Magnusson and Schaefer)
1f:	0.45	Pople and co-workers recommended value
	0.51	Optimal valence SD-CI value (Feller)
2f:	(0.24, 0.66)	Optimal valence SD-CI value (Feller)
Sulfur		
1d:	0.65	6-31G**, average of optimal SCF values
5d:	(0.08, 0.23, 0.67, 2.12, 8.35)	Optimal SCF for S$^+$, 3s^{2}3p^{2}3d (^{4}F)
1f:	0.55	Pople and co-workers recommended value
Chlorine		
1d:	0.75	6-31G**, average of optimal SCF values
5d:	(0.085, 0.24, 0.71, 2.25, 8.90)	Recommended SCF values (Magnusson and Schaefer)
1f:	0.70	Pople and co-workers recommended value
Argon		
1d:	0.85	6-31G**, average of optimal SCF values

[a] References (19) and (20).

[b] D. Feller (unpublished).

[c] T. H. Dunning, Jr.[99] Similar exponents have also been given by Feller and Davidson.[100]

[d] Frisch et al.[61]

Certain molecular states and properties require additional basis set flexibility beyond what is normally available from sets derived from strictly energetic considerations. For instance, in the case of negatively charged species, the additional electron is often found in a more diffuse orbital than is normal. The usual valence basis must be augmented if even qualitative agreement with experiment is expected. The 3-21+G and 3-21++G basis sets[118] possess an

Table 7 MR SD-CI Contributions to the Correlation Energy from Functions of Increasing *L* Value[a]

	s	p	d	f	Total
H_2	−0.0264 (65%)	−0.0124 (30%)	−0.0016 (4%)	−0.0003 (1%)	−0.0407
	s, p	d	f	g	Total[b]
N (^{4}S)	−0.0936 (51%)	−0.0762 (42%)	−0.0106 (6%)	−0.0019 (1%)	−0.1823
N^+ (^{3}P)	−0.0929 (58%)	−0.0608 (38%)	−0.0061 (4%)	−0.0013 (1%)	−0.1611
O (^{3}P)	−0.1261 (51%)	−0.0964 (39%)	−0.0209 (8%)	−0.0047 (2%)	−0.2481
O^- (^{2}P)	−0.1822 (57%)	−0.1021 (32%)	−0.0250 (8%)	−0.0084 (3%)	−0.3177
F (^{2}P)	−0.1582 (51%)	−0.1148 (37%)	−0.0293 (10%)	−0.0058 (2%)	−0.3079

[a] Energies in hartrees, with percentages of the total given in parentheses. The energy contributions for various groups of functions came from uncontracted even-tempered Gaussian calculations with basis sets as large as (12s, 6p, 5d, 4f) on H_2 (E_{CI} = −1.1743) and (23s, 12p, 10d, 4f, 2g) on the first-row atoms.

[b] The total computed correlation energy. For comparison purposes, the empirical estimates of the correlation energy for these states are −0.0409 H_2 ($^1\Sigma_g$), −0.1889 N (^{4}S), −0.1735 N^+ (^{3}P), −0.2596 O (^{3}P), −0.3329 O^- (^{2}P), −0.3253 F (^{2}P).

extra (s,p) diffuse set of functions on elements Li–F and an extra s set on hydrogen, respectively, to improve the description of anion geometries.

Rydberg states are characterized by an electron occupying a very diffuse orbital. Even large energy-optimized basis sets will not contain sufficiently small exponents to handle such states. Suggested Rydberg state (s,p,d) exponents for B–F are given by Dunning and Hay.[52]

The isotropic hyperfine coupling constant (A_{iso}) is a measure of the unpaired spin density at a particular nucleus. Experience has shown that accurate calculations of this property for first-row atoms require large (s,p,d,f) basis sets with enough flexibility in the core to adequately reproduce the large negative core contribution to A_{iso} that will nearly cancel the positive valence contribution.[119] A very wide range of answers is possible by unintentionally shifting the basis set in the direction of the core or valence regions. The general notion of a "balanced" basis set was introduced into the basis set literature by Mulliken[120] in a slightly different context. He warned of the need for "balanced" basis sets for molecules containing different elements, so as to avoid unphysical build-ups of charge on any one species.

Weak Interactions

Attempts to compute by *ab initio* methods the interaction energy of two systems that are only weakly bound, such as hydrogen-bonded water molecules or van der Waals complexes, have proven exceedingly difficult. The origin of the difficulty primarily stems from the small magnitude of the quantity being sought relative to the basis set superposition error discussed previously. For example, van Lenthe et al.[121] recently reported a potential curve for the He_2 van der Waals complex. Using a very large (15s,4p,4d,2f,1g,1h) → [7s,4p,4d,2f,1g,1h] contracted Gaussian basis and a sophisticated MCSCF/MR SD-CI wavefunction, they obtained a well depth of 0.03439 mhartrees (0.022 kcal/mol), in good agreement with experimental measurements only after correcting for the BSSE of 0.03978 mhartrees.

With this as an introduction, it may be wondered if weak interactions are presently within the reach of *ab initio* methods. The answer, according to a recent review by Chalasinski and Gutowski,[122] appears to be a qualified "yes." Although the BSSE at the correlated wave function level is perhaps beyond our means to entirely eliminate by attempting to saturate the basis, the counterpoise correction scheme seems to work well. For dimers such as He_2 and $(H_2)_2$, where the attraction comes almost entirely from what is known as the "dispersion" force, rather high L values are needed. For many others, such as Li^+He, the demand placed on the basis set is not as great.

For hydrogen-bonded structures the situation is a little better. In an SCF study of the BSSE by Schwenke and Truhlar,[123,124] the authors focused on the interaction energy of HF + HF using basis sets, which ranged from STO-3G to an extended basis that included multiple sets of d's on fluorine and p's on hydrogen plus bond functions. Regarding the effectiveness of the full counterpoise correction, the authors conclude "that counterpoise corrections do not systematically improve the accuracy obtained with small basis sets. Since we also find the counterpoise corrections do not become more reliable as the basis set size is increased, we conclude that in general the extra expense of a counterpoise correction is not warranted." The magnitude of the BSSE fluctuated between 11 kcal/mol (STO-3G) and 0.3 kcal/mol (extended) out of a total interaction energy of 19 kcal/mol.

Even larger basis sets were employed by Frisch et al.[125] in a follow-up study that included correlation recovery at the MP4 level and reported excellent agreement with extensive experimental data. Here again, the authors concluded that "Counterpoise estimates of the basis set superposition error do not provide quantitative information about basis set deficiencies in studies of hydrogen bonded complexes and are a poor substitute for expansion of the basis set if quantitative accuracy is desired."

Taking the opposite view, Szalewicz et al.[126] examined the water dimer with still larger basis sets involving 212 (s,p,d,f) functions. They found that the SCF interaction energy could be satisfactorily reproduced with a small double-zeta quality basis plus properly chosen polarization functions, when used in conjunction with the counterpoise method. The correlation contribution to the interaction energy was computed to be -1.0 ± 0.3 kcal/mol. Alberts et al.[127] have also recently advocated the use of the full counterpoise method for weakly bound complexes of HF and CO, CO_2 and N_2CO.

Dykstra and co-workers have argued that the counterpoise method should be used only as a probe of the qualitative magnitude of the BSSE, because it sometimes overestimates and sometimes underestimates the BSSE. They proposed a modified counterpoise method for use in hydrogen-bonding situations, which they call the "polarization" counterpoise method.[128] The new method is based on the premise that a more accurate measure of the BSSE can be obtained from measuring the energy lowering upon bringing monomer A up to the functions on center B if A experiences the electric field of B as it would in the actual dimer. Because the conventional counterpoise method measures the energy lowering for an isolated monomer, it will most likely underestimate the BSSE associated with hydrogen bonding. In Dykstra's approach the field of the other monomer is represented by appropriately placed positive and negative charges equal to the nuclear and electronic charge on the other monomer.

HF is a highly polar molecule. Dykstra[129] has shown that a classical electrostatic approach using polarizable charge distributions can often accurately reproduce the results of high level *ab initio* calculations. The advantage of this approach is that it requires high quality calculations on the monomer only. This is to obtain accurate moments (e.g., dipole or quadrupole), polarizabilities, and hyperpolarizability. Very extended basis sets are needed.

If μ is the electric dipole moment of a molecule, then in the presence of an external electric field E, μ can be expressed as

$$\mu = \mu^0 + \alpha E + \kappa\beta E^2 + \lambda\gamma E^3 + \cdots$$

where μ^0 is the permanent dipole moment, α is the first order (dipole) polarizability, β is the second order hyperpolarizability, etc. The symbols κ, λ . . . are fixed numerical constants chosen by convention. (A more detailed description of this approach is described in the chapter by Dykstra et al.) Work by Purvis and Bartlett on H_2O[130] and Sekino and Bartlett on HF[131] shows the need for quite diffuse functions. For water, center-of-mass basis functions with exponents as small as 0.0079 for s and 0.0057 for p were employed along with a large [7s,5p,3d/5s,3p] basis set on the atoms. For HF an even larger [5s,3p,4d,2f/5s,3p] basis, capable of yielding 92% of the valence correlation energy at the MP4 level, was used. In spite of the number of basis functions, the β hyperpolarizability was still only 50–70% of the experimental value in water.

CONCLUSION

As long as molecular wavefunctions are expanded in terms of finite basis sets, there will continue to be a need to calibrate the accuracy of these sets with regard to the ever-increasing list of properties they are asked to provide.

For the sake of the nonspecialist, this information must be both readily available and readily digestible. The *ab initio* model suffers from a well-deserved reputation for being computationally expensive. However, as the trend toward improved hardware and software technology accelerates, the *ab initio* model is employed in more and more laboratories as a useful adjunct to experimental techniques. The same is true of the other computational chemistry methods. Researchers must learn where each computational procedure can be used with good assurance of success and where to treat the computer's answer with skepticism.

REFERENCES

1. See for example, D. Sundholm, P. Pyykkö, and L. Laaksonen, *Mol. Phys.*, *56*, 1411 (1985). Two-Dimensional, Fully Numerical Molecular Calculations. 10. Hartree–Fock Results for He_2, Li_2, Be_2, HF, OH, N_2, CO, BF, No^+ and CN^-.
2. U. Burkert and N. L. Allinger, *Molecular Mechanics*. ACS Monograph Series, Washington, D.C., 1982.
3. M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, *99*, 4899 (1977). Ground States of Molecules. 38. The MNDO Method.
4. D. N. Nanda and K. Jug, *Theor. Chim. Acta*, *57*, 95 (1980). SINDO1. A Semiempirical SCF MO Method for Molecular Binding Energy and Geometry.
5. E. R. Davidson and D. Feller, *Chem. Rev.*, **86**, 681 (1986). Basis Set Selection for Molecular Calculations.
6. S. Huzinaga, *Comput. Phys. Rep.*, **2**, 279 (1985). Basis Sets for Molecular Calculations.
7. R. Ahlrichs and P. R. Taylor, *J. Chim. Phys.*, **78**, 315 (1981). The Choice of Gaussian Basis Sets for Molecular Electronic Structure Calculations.
8. S. Wilson, in *Methods in Computational Molecular Physics*, G. H. Diercksen and S. Wilson, Eds. Reidel, Dordrecht. The Netherlands, 1983, pp. 71–93. Basis Sets.
9. J. C. Slater, *Phys. Rev.*, *36*, *57* (1930). Atomic Shielding Constants.
10. See for example, M. Sekiya and H. Tatewaki, *J. Chem. Phys.*, **86**, 2891 (1987). Accurate STF HF Wave Functions for B to Ca and Accurate STF HF Wave Functions for N_2 and P_2; or M. R. A. Blomberg and B. Liu, *J. Chem. Phys.*, **89**, 6870 (1988). A Systematic Approach to Transition Moment Calculations.
11. S. F. Boys, *Proc. R. Soc. (London) Ser. A*, **200**, 542 (1950). Electronic Wave Functions. II. A Calculation for the Ground State of the Beryllium Atom.
12. A. A. Frost, *J. Chem. Phys.*, **47**, 3707 (1967). Floating Spherical Gaussian Orbital Model of Molecular Structure. I. Computational Procedure. LiH as an Example; and A. A. Frost, *J. Chem. Phys.*, **47**, 3714 (1967). Floating Spherical Gaussian Orbital Model of Molecular Structure. II. One- and Two-Electron-Pair Systems.
13. P. T. van Duijnen and D. B. Cook, *Mol. Phys.*, **21**, 475 (1971). Ab Initio Calculations with Small Ellipsoidal Gaussian Basis Sets.
14. J. L. Whitten, *J. Chem. Phys.*, **44**, 359 (1966). Gaussian Lobe Function Expansions of Hartree–Fock Solutions for First-Row Atoms and Ethylene.
15. C. M. Reeves and M. C. Harrison, *J. Chem. Phys.*, **39**, 11 (1963). Use of Gaussian Functions in the Calculation of Wavefunctions for Small Molecules. II. The Ammonia Molecule.

16. E. Clementi and D. R. Davis, *J. Comput. Phys.*, 1, 223 (1966). Electronic Structure of Large Molecular Systems.

17. R. C. Raffenetti, *J. Chem. Phys.*, 58, 4452 (1973). General Contraction of Gaussian Atomic Orbitals: Core, Valence, Polarization and Diffuse Basis Sets; Molecular Integral Evaluation.

18. R. D. Bardo and K. Ruedenberg, *J. Chem. Phys.*, 60, 918 (1974). Systematic Approach to Extended Even-tempered Orbital Bases for Atomic and Molecular Calculations.

19. P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, 28, 213 (1973). The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies.

20. M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, *J. Chem. Phys.*, 77, 3654 (1982). Self-Consistent Molecular Orbital Methods. XXIII. A Polarization-Type Basis Set for Second-Row Elements.

21. B. J. Rosenberg and I. Shavitt, *J. Chem. Phys.*, 63, 2162 (1975). *Ab Initio* SCF and CI Studies on the Ground State of the Water Molecule. I. Comparison of CGTO and STO Basis Sets Near the Hartree–Fock Limit.

22. P. E. Regier, J. Fisher, B. S. Sharma, and A. J. Thakkar, *Int. J. Quantum Chem.*, 28, 429 (1985). Gaussian vs. Slater Representations of d Orbitals: An Information Theoretic Appraisal Based on Both Position and Momentum Space Properties.

23. P. Carsky and M. Urban, *Ab Initio Calculations*. Springer-Verlag, Heidelberg, 1980.

24. K. McDowell, *J. Chem. Phys.*, 68, 4151 (1978). The Incomplete Basis Set Problem. I. Perturbative Corrections to Hartree–Fock Energies.

25. B. Klahn, *J. Chem. Phys.*, 83, 5754 (1985). The Convergence of CI Calculations for Atomic and Molecular Electronic Bound States in a Basis of Floating Gaussian Orbitals.

26. See for example, D. Feller and E. R. Davidson, *J. Chem. Phys.*, 88, 7580 (1988). A Multireference CI Determination of the Isotropic Hyperfine Constants for First Row Atoms B-F.

27. S. F. Boys and F. Bernardi, *Mol. Phys.*, 19, 553 (1970). The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors.

28. A. Johnson, P. Kollman, and S. Rothenberg, *Theor. Chim. Acta*, 29, 167 (1973). An Application of the Functional Boys-Bernardi Counterpoise Method to Molecular Potential Surfaces.

29. K. Morokuma and K. Kitaura, *Chemical Applications of Atomic and Molecular Electrostatic Potentials*. P. Politzer, Ed., Plenum, New York, 1981.

30. M. Gutowski, F. B. van Duijneveldt, G. Chalasinski, and L. Piela, *Mol. Phys.*, 61, 233 (1987). Proper Correction for the Basis Set Superposition Error in SCF Calculations of Intermolecular Interactions.

31. M. Gutowski, J. H. van Lenthe, J. Verbeek, and F. B. van Duijneveldt, *Chem. Phys. Lett.*, 124, 370 (1986). The Basis Set Superposition Error in Correlated Electronic Structure Calculations.

32. J. Andzelm, M. Klobukowski, and E. Radzio-Andzelm, *J. Comput. Chem.*, 5, 146 (1984). Compact Contracted Gaussian-Type Basis Sets for Halogen Atoms. Basis Set Superposition Effects on Molecular Properties.

33. GAUSSIAN 86, M. J. Frisch, J. S. Binkley, H. B. Schlegel, K. Raghavachari, C. F. Melius, R. L. Martin, J. J. P. Stewart, F. B. Bobrowicz, C. M. Rohlfing, L. R. Kahn, D. J. DeFrees, R. Seeger, R. A. Whiteside, D. J. Fox, E. M. Fleuder, and J. A. Pople, Carnegie-Mellon Quantum Chemistry Publishing Unit, Pittsburgh, PA, 1984.

34. HONDO version 7.0, M. Dupuis, J. D. Watts, H. O. Villar, and G. J. B. Hurst, 1987.

35. M. Dupuis, D. Spangler, and J. J. Wendoloski, NRCC Software Catalog, Vol. 1 Program QG01 (GAMESS), 1980. This program was heavily modified by M. W. Schmidt and S. T. Elbert.

36. D. Feller, C. M. Boyle, and E. R. Davidson, *J. Chem. Phys.*, 86, 3424 (1987). One-Electron Properties of Several Small Molecules Using Near Hartree–Fock Limit Basis Sets.

37. A. N. Tavouktsoglou and S. Huzinaga, *J. Chem. Phys.*, 72, 1385 (1980). A New Minimal Gaussian Basis Set for Molecular Calculations.

38. W. J. Hehre, L. Radom, P. v.R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*. Wiley, New York, 1986.

39. W. J. Hehre, R. Stewart, and J. A. Pople, *J. Chem. Phys.*, 51, 2657 (1969). Self-Consistent Molecular Orbital Methods 1. Use of Gaussian Expansions of Slater-Type Atomic Orbitals.

40. W. J. Hehre, R. Ditchfield, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, 52, 2769 (1970). Self-Consistent Molecular Orbital Methods. 4. Use of Gaussian Expansions of Slater-type Orbitals. Extensions to Second Row Molecules.

41. W. J. Pietro, B. A. Levi, W. J. Hehre, and R. F. Stewart, *Inorg. Chem.*, 19, 2225 (1980). Molecular Orbital Theory of the Properties of Inorganic and Organometallic Compounds. 1. STO-NG Basis Sets for Third Row Main-Group Elements.

42. W. J. Pietro, E. S. Blurock, R. S. Hout, W. J. Hehre, D. J. DeFrees, and R. F. Stewart, *Inorg. Chem.*, 20, 3650 (1981). Molecular Orbital Theory of the Properties of Inorganic and Organometallic Compounds. 1. STO-NG Basis Sets for Fourth-Row Main Group Elements.

43. W. J. Pietro and W. J. Hehre, *J. Comput. Chem.*, 4, 241 (1983). Molecular Orbital Theory of the Properties of Inorganic and Organometallic Compounds. 1. STO-3G Basis Sets for First- and Second-Row Transition Metals.

44. J. S. Binkley, J. A. Pople, and W. J. Hehre, *J. Am. Chem. Soc.*, 102, 939 (1980). Self-Consistent Molecular Orbital Methods. 21. Small Split-Valence Basis Sets for First Row Elements.

45. M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro, and W. J. Hehre, *J. Am. Chem. Soc.*, 104, 2797 (1982). Self-Consistent Molecular Orbital Methods. 21. Small Split-Valence Basis Sets for Second Row Elements.

46. W. J. Pietro, M. M. Francl, W. J. Hehre, J. A. Pople, and J. S. Binkley, *J. Am. Chem. Soc.*, 104, 5048 (1982). Self-Consistent Molecular Orbital Methods. 24. Supplemented Small Split-Valence Basis Sets for Second-Row Elements.

47. R. Ditchfield, W. J. Hehre, and J. A. Pople, *J. Chem. Phys.*, 54, 724 (1971). Self-Consistent Molecular Orbital Methods 9. An Extended Gaussian-Type Basis for Molecular Orbital Studies of Organic Molecules.

48. W. J. Hehre and W. A. Lathan, *J. Chem. Phys.*, 56, 5255 (1972). Self-Consistent Molecular Orbital Methods 14. An Extended Gaussian-Type Basis for Molecular Orbital Studies of Organic Molecules. Inclusion of Second Row Elements.

49. W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.*, 56, 2257 (1972). Self-Consistent Molecular Orbital Methods 12. Further Extension of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules.

50. P. Pulay, G. Fogarasi, F. Pang, and J. E. Boggs, *J. Am. Chem. Soc.*, 101, 2550 (1979). Systematic *Ab Initio* Gradient Calculations of Molecular Geometries, Force Constants and Dipole Moment Derivatives.

51. R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, *J. Chem. Phys.*, 72, 650 (1980). Self-Consistent Molecular Orbital Methods. XX. A Basis Set for Correlated Wave Functions.

52. T. H. Dunning, Jr., and P. J. Hay, *Modern Theoretical Chemistry*, Vol. 2. H. F. Schaefer, Ed., Plenum Press, New York, 1977.

53. J. Andzelm, M. Klobukowsi, E. Radzio-Andzelm, Y. Saski, and H. Tatewaki, *Gaussian Basis Sets for Molecular Calculations*. S. Huzinaga, Ed., Elsevier, Amsterdam, 1984.

54. T. H. Dunning, Jr., *J. Chem. Phys.*, 53, 2823 (1970). Gaussian Basis Functions for Use in Molecular Calculations. I. Contraction of (9s,5p) Atomic Basis Sets for the First-Row Atoms.

55. J. A. Pople, H. B. Schlegel, R. Krishnan, D. J. DeFrees, J. S. Binkley, M. J. Frisch, R. A. Whiteside, R. F. Hout, and W. J. Hehre, *Int. J. Quantum Chem. Symp.*, 15, 269 (1981). Molecular Orbital Studies of Vibrational Frequencies.

56. R. F. Hout, B. A. Levi, and W. J. Hehre, *J. Comput. Chem.*, 3, 234 (1982). Effects of Electron Correlation on Theoretical Vibrational Frequencies.

57. H. B. Schlegel, S. Wolfe, and F. Bernardi, *J. Chem. Phys.*, 67, 4181 (1977). *Ab Initio*

Computation of Force Constants. IV. The Theoretical Anharmonic Force Fields and Vibrational Frequencies of Methylamine, Methanol and Methanethiol.

58. D. F. McMillen and D. M. Golden, *Annu. Rev. Phys. Chem.*, 33, 493 (1982). Hydrocarbon Bond Dissociation Energies.

59. E. R. Davidson, *The World of Quantum Chemistry*. R. Daudel and B. Pullman, Eds., Reidel, Dordrecht, 1974.

60. See for example, P. Saxe, H. F. Schaefer III and N. Handy, *Chem. Phys. Lett.*, 79, 202 (1981). Exact Solution (within a Double Zeta Basis Set) of the Schrödinger Electronic Equation for Water; or C. W. Bauschlicher, Jr. and S. R. Langhoff, *J. Chem. Phys.*, 86, 5595 (1987). Full CI Benchmark Calculations on N_2, NO and O_2: A Comparison of Methods for Describing Multiple Bonds.

61. M. J. Frisch, J. A. Pople, and J. S. Binkley, *J. Chem. Phys.*, 80, 3265 (1984). Self-Consistent Molecular Orbital Methods. 25. Supplementary Functions for Gaussian Basis Sets.

62. S. Huzinaga, *J. Chem. Phys.*, 42, 1293 (1965). Gaussian-Type Functions for Polyatomic Systems.

63. T. H. Dunning, *J. Chem. Phys.*, 55, 716 (1971). Gaussian Basis Functions for Use in Molecular Calculations. III. Contraction of (10s,6p) Atomic Basis Sets for First Row Atoms.

64. F. B. van Duijneveldt, *IBM Tech. Res. Rep.* RJ945 (1971). Gaussian Basis Sets for Atoms H–Ne for Use in Molecular Calculations.

65. K. Faegri and J. Almlöf, *J. Comput. Chem.*, 7, 396 (1986). Energy-Optimized GTO Basis Sets for LCAO Calculations. A Gradient Approach.

66. H. Partridge, *J. Chem. Phys.*, 87, 6643 (1987). Near Hartree–Fock Quality GTO Basis Sets for Second Row Atoms.

67. H. Partridge, *J. Chem. Phys.*, 90, 1043 (1989). Near Hartree–Fock Quality GTO Basis Sets for First- and Third-Row Atoms.

68. K. Ruedenberg, R. C. Raffenetti, and R. D. Bardo, in *Energy, Structure and Reactivity, Proceedings of the 1972 Boulder Conference*. Wiley, New York, 1973.

69. C. M. Reeves, *J. Chem. Phys.*, 39, 1 (1963). Use of Gaussian Functions in the Calculation of Wavefunctions for Small Molecules. I. Preliminary Investigations.

70. R. C. Raffenetti, *J. Chem. Phys.*, 59, 5936 (1973). Even-Tempered Atomic Orbitals. II. Atomic SCF Wavefunctions in Terms of Even-Tempered Exponential Bases.

71. R. D. Bardo and K. Ruedenberg, *J. Chem. Phys.*, 59, 5956 (1974). Even-Tempered Atomic Orbitals. III. Economic Deployment of Gaussian Primitives in Expanding Atomic SCF Orbitals.

72. R. C. Raffenetti, *Int. J. Quantum. Chem. Symp.*, 9, 289 (1975). Even-Tempered Gaussian Atomic Orbitals: First-Row Atoms.

73. D. Feller and K. Ruedenberg, *Theor. Chim. Acta*, 71, 23 (1979) 1. Systematic Approach to Extended Even-Tempered Orbital Bases for Atomic and Molecular Calculations.

74. M. W. Schmidt and K. Ruedenberg, *J. Chem. Phys.*, 71, 3951 (1979). Effective Convergence to Complete Orbital Bases and to the Atomic Hartree–Fock Limit Through Systematic Sequences of Gaussian Primitives.

75. S. Wilson, *Theor. Chim. Acta*, 58, 31 (1980). Universal Systematic Sequences of Even-Tempered Gaussian Primitive Functions in Electronic Correlation Studies.

76. S. Wilson, *Theor. Chim. Acta*, 57, 53 (1980). Systematic Sequences of Even-Tempered Gaussian Primitive Functions in Electronic Correlation Studies Using Many-Body Perturbation Theory.

77. E. Clementi and G. Corongiu, *Chem. Phys. Lett.*, 90, 359 (1982). Geometrical Basis Set for Molecular Computations.

78. S. Huzinaga, M. Klobukowski, and H. Tatewaki, *Can. J. Chem.*, 63, 1812 (1985). The Well-Tempered GTF Basis Sets and Their Applications in the SCF Calculations on N_2, CO, Na_2 and P_2.

79. S. Huzinaga and M. Klobukowski, *Chem. Phys. Lett.*, 120, 509 (1985). The Well-Tempered GTF Basis Sets for K Through Xe.

80. M. Klobukowski and S. Huzinaga, *J. Chem. Phys.*, 84, 323 (1986). The Well-Tempered GTF Basis Sets in Calculations on the s^2d^n, s^1d^{n+1} and d^{n+2} Configurations of the First- and Second-Row Transition Metal Atoms.

81. R. Poirier, R. Kari, and I. G. Csizmadia, *Handbook of Gaussian Basis Sets.* Elsevier Science Publishing, New York, 1985.

82. S. Huzinaga and Y. Sakai, *J. Chem. Phys.*, 50, 1371 (1969). Gaussian-Type Functions for Polyatomic Systems.

83. H. Tatewaki and S. Huzinaga, *J. Comput. Chem.*, 3, 205 (1980). A Systematic Preparation of New Contracted Gaussian-Type Orbital Sets. III. Second-Row Atoms from Li Through Ne.

84. H. Tatewaki, *J. Comput. Chem.*, 6, 237 (1985). A Systematic Preparation of New Contracted Gaussian-Type Orbitals. IX. [54/5], [64/5], [64/6], [74/6], [74/7] and MAXI-1–MAXI-5 from Li to Ne.

85. H. Tatewaki and S. Huzinaga, *J. Chem. Phys.*, 71, 4339 (1979). A Systematic Preparation of New Contracted Gaussian-Type Orbital Sets. I. Transition Metal Atoms from Sc to Zn.

86. H. Tatewaki and S. Huzinaga, *J. Chem. Phys.*, 72, 399 (1980). A Systematic Preparation of New Contracted Gaussian-Type Orbital Basis Sets. II. Test Basis Set for Cu_2 Molecule with and without Splitting of the Outer Orbitals.

87. H. Tatewaki, Y. Sakai, and S. Huzinaga, *J. Comput. Chem.*, 2, 96 (1981). A Systematic Preparation of New Contracted Gaussian-Type Orbital Sets. IV. The Effect of Additional 3s Functions Introduced by the Use of the Six-Membered 3d GTO's.

88. H. Tatewaki, Y. Sakai, and S. Huzinaga, *J. Comput. Chem.*, 2, 278 (1981). A Systematic Preparation of New Contracted Gaussian-Type Orbital Sets. VII. MINI-3, MINI-4, MIDI-3 and MIDI-4 Sets for Transition Metal Atoms.

89. Y. Sakai, H. Tatewaki, and S. Huzinaga, *J. Comput. Chem.*, 3, 6 (1982). A Systematic Preparation of New Contracted Gaussian-Type Orbital Sets. VIII. MINI-1 and MIDI-1 Sets for Ga Through Cd.

90. L. Seijo, Z. Barandiarian, M. Klobukowski, and S. Huzinaga, *Chem. Phys. Lett.*, 117, 151 (1985). *Ab Initio* Calculations on Transition Metal Compounds Using Small Minimal GTO Basis Sets.

91. A. D. McLean and G. S. Chandler, *J. Chem. Phys.*, 72, 5639 (1980). Contracted Gaussian Basis Sets for Molecular Calculations. I. Second Row Atoms, $Z = 11–18$.

92. Examples of programs that do handle the general contraction scheme efficiently are BIGGMOLLI by R. C. Raffenetti, *QCPE Catalog* 11, 328 (1977) and ARGOS by R. M. Pitzer.

93. J. Almlöf and P. R. Taylor, *J. Chem. Phys.*, 86, 4070 (1987). General Contraction of Gaussian Basis Sets. I. Atomic Natural Orbitals for First- and Second-Row Atoms.

94. K. Jankowski, R. Becherer, P. Scharf, H. Schiffer, and R. Ahlrichs, *J. Chem. Phys.*, 82, 1413 (1985). The Impact of Higher Polarization Basis Functions on Molecular *Ab Initio* Results. I. The Ground State of F_2.

95. J. Almlöf, T. Helgaker, and P. R. Taylor, *J. Phys. Chem.*, 92, 3029 (1988). Gaussian Basis Sets for High-Quality *Ab Initio* Calculations.

96. T. Helgaker and J. Almlöf, *J. Chem. Phys.*, 89, 4889 (1988). Molecular Wave Functions and Properties Calculated Using Floating Gaussian Orbitals.

97. See for example, A. O. Bawagan, C. E. Brion, E. R. Davidson, and D. Feller, *Chem. Phys.*, 113, 19 (1987). Electron Momentum Spectroscopy of the Outer Valence Orbitals of H_2O and D_2O.

98. C. W. Bauschlicher and S. R. Langhoff, *Chem. Phys. Lett.*, 145, 205 (1988). Theoretical Confirmation of a $^4\Delta$ Ground State of FeH.

99. T. H. Dunning, Jr., *J. Chem. Phys.*, 90, 1007 (1989). Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen.

100. D. Feller and E. R. Davidson, *J. Chem. Phys.*, 82, 4135 (1985). *Ab Initio* Multireference CI Determinations of the Electron Affinity of Carbon and Oxygen.

101. H. Huber, *Chem. Phys. Lett.*, 62, 95 (1979). Geometry Optimization in *Ab Initio* SCF Calculation. Floating Orbital Geometry Optimization Applying the Hellmann–Feynman Force.

102. L. Adamowicz and R. J. Bartlett, *Chem. Phys. Lett.*, 105, 167 (1984). Extended Floating Spherical Gaussian Basis Sets for Molecules. Generation Procedure and Results for H_2O.

103. B. Roos and P. Siegbahn, *Theor. Chim. Acta*, 17, 209 (1970). Gaussian Basis Sets for the First and Second Row Atoms.

104. K. Faegri and H. J. Speis, *J. Chem. Phys.*, 86, 7035 (1987). Basis Set Quality vs. Size. Approximate Gaussian-Type Orbitals (GTO) Wavefunctions for First Row Transition Metal Atoms.

105. H. Basch, C. J. Hornback, and J. W. Moskowitz, *J. Chem. Phys.*, 51, 1311 (1969). Gaussian-Orbital Basis Sets for First Row Transition-Metal Atoms.

106. A. J. H. Wachters, *J. Chem. Phys.*, 52, 1033 (1970). Gaussian Basis Sets for Molecular Wavefunctions Containing Third-Row Atoms.

107. B. Roos, A. Veillard, and G. Vinot, *Theor. Chim. Acta*, 20, 1 (1971). Gaussian Basis Sets for Molecular Wavefunctions Containing Third-Row Atoms.

108. P. J. Hay, *J. Chem. Phys.*, 66, 4377 (1977). Gaussian Basis Sets for Molecular Calculations. The Representation of 3d Orbitals in Transition Metal Atoms.

109. S. Rothenberg and H. F. Schaeffer III, *J. Chem. Phys.*, 54, 2765 (1971). Methane as a Numerical Experiment for Polarization Basis Function Selection.

110. W. Butscher, S. Shih, R. J. Buenker, and S. D. Peyerimhoff, *Chem. Phys. Lett.*, 52, 457 (1977). Configuration Interaction Calculation for the N_2 Molecule and Its Three Lowest Dissociation Limits.

111. D. Neisius and G. Verhaegen, *Chem. Phys. Lett.*, 66, 358 (1979). Bond Functions for *Ab Initio* Calculations on Polyatomic Molecules. Hydrocarbons.

112. L. Lawlor, K. Vasudevan, and F. Grein, *Chem. Phys. Lett.*, 75, 79 (1980). Bond Functions and 3d Polarization Functions for Fluorine Bonded Molecules.

113. J. S. Wright and R. J. Williams, *J. Chem. Phys.*, 78, 5264 (1983). Accurate *Ab Initio* Potential Curves Using Bond Functions.

114. J. S. Wright and R. J. Williams, *J. Chem. Phys.*, 79, 2893 (1983). MRD-CI Potential Surfaces Using Balanced Basis Sets. I. First-Row Diatomic Hydrides.

115. J. S. Wright and R. J. Buenker, *Chem. Phys. Lett.*, 106, 570 (1984). The Effect of Bond Functions on Molecular Dissociation Energies.

116. C. W. Bauschlicher, *Chem. Phys. Lett.*, 122, 572 (1985). The Effect of Bond Functions on Dissociation Energies.

117. P. Mach and O. Kysel, *J. Comput. Chem.*, 6, 312 (1985). Comparison Between 6-31G + Bond Functions and 6-31G* Basis Sets in UMP2 Calculations of Dissociation Energies of the First Row Element Compounds.

118. T. Clark, J. Chandrasekhar, G. Spitznagel, and P. von Rague Schleyer, *J. Comput. Chem.*, 4, 294 (1983). Efficient Diffuse Function-Augmented Basis Sets for Anion Calculations. III. The 3-21G+G Basis Set for First-Row Elements, Li–F.

119. See for example, C. W. Bauschlicher, S. R. Langhoff, and H. Partridge, *J. Chem. Phys.*, 89, 2985 (1988). Theoretical Study of the Nitrogen Atom Hyperfine Coupling Constant; or L. B. Knight, J. T. Petty, S. T. Cobranchi, D. Feller, and E. R. Davidson, *J. Chem. Phys.*, 88, 3441 (1988). The Generation of $^{12}C^{31}P$ and $^{13}C^{31}P$ by Reactive Laser Vaporization for Rare Gas Matrix ESR Studies: Comparison with *Ab Initio* Theoretical Calculations.

120. R. S. Mulliken, *J. Chem. Phys.*, 36, 3428 (1962). Criteria for the Construction of Good Self-Consistent-Field Molecular Orbital Wave Functions, and the Significance of the LCAO-MO Population Analysis.

121. J. H. van Lenthe, R. J. Vos, J. G. van Duijneveldt-van de Rijdt, and F. B. van Duijneveldt,

Chem. Phys. Lett., 143, 435 (1988). An Accurate *Ab Initio* Calculation of the He_2 Potential Curve.

122. G. Chalasinski and M. Gutowski, *Chem. Rev.*, 88, 943 (1988). Weak Interactions between Small Systems. Models for Studying the Nature of Intermolecular Forces and Challenging Problems for *Ab Initio* Calculations.
123. D. W. Schwenke and D. G. Truhlar, *J. Chem. Phys.*, 82, 2418 (1985). Systematic Study of Basis Set Superposition Errors in the Calculated Interaction Energy of Two HF Molecules.
124. D. W. Schwenke and D. G. Truhlar, *J. Chem. Phys.*, 84, 4113 (1985). Erratum: Systematic Study of Basis Set Superposition Errors in the Calculated Interaction Energy of Two HF Molecules.
125. M. Frisch, J. E. Del Bene, J. S. Binkley, and H. F. Schaefer III, *J. Chem. Phys.*, 84, 2279 (1986). Extensive Theoretical Studies of the Hydrogen-Bonded Complexes $(H_2O)_2$, $(H_2O)_2H^+$, $(HF)_2$, $(HF)_2H^+$, F_2H^-, and $(NH_3)_2$.
126. K. Szalewicz, S. J. Cole, W. Kolos, and J. J. Bartlett, *J. Chem. Phys.*, 89, 3662 (1988). A Theoretical Study of the Water Dimer Interaction.
127. I. L. Alberts, N. C. Handy, and E. D. Simandiras, *Theor. Chim. Acta*, 74, 415 (1988). The Structure and Harmonic Vibrational Frequencies of the Weakly Bound Complexes Formed by HF with CO, CO_2 and N_2CO.
128. S. K. Loushin, S. Liu, and C. E. Dykstra, *J. Chem. Phys.*, 84, 2720 (1986). Improved Counterpoise Corrections for the *Ab Initio* Calculation of Hydrogen Bonding Interactions.
129. See for example, C. E. Dykstra, *Acc. Chem. Res.*, 21, 355 (1988). Intermolecular Electrical Interaction: A Key Ingredient in Hydrogen Bonding.
130. G. D. Purvis and R. J. Bartlett, *Phys. Rev. A*, 23, 1594 (1981). Molecular Hyperpolarizabilities. II. A Correlated Study of H_2O.
131. H. Sekino and R. J. Bartlett, *J. Chem. Phys.*, 84, 2726 (1986). Hyperpolarizabilities of the Hydrogen Fluoride Molecule: A Discrepancy Between Theory and Experiment?.