Chem. Rev. 1986, 86, 681–696

Basis Set Selection for Molecular Calculations

ERNEST R. DAVIDSON* and DAVID FELLER

Indiana University, Bloomington, Indiana 47405

Received January 10, 1986 (Revised Manuscript Received April 5, 1986)

Introduction

The quantum chemistry literature contains references to a plethora of basis sets, currently numbering almost 100. While professional quantum chemists might become familiar with several dozen of these in a lifetime of calculations, the occasional user of ab initio programs probably wishes to ignore all but the two or three sets which, through habitual use, have become personal favorites. Unfortunately, this attitude has its drawbacks. Intelligent reading of the literature requires at least a cursory knowledge of the limitations of other basis sets. Information concerning the likely accuracy of a specific basis for a particular property is essential in order to judge the adequacy of the computational method and, hence, the soundness of the results. Occasionally, for reasons of economy or computational feasibility, a basis set is selected for which the computed results are nearly without significance.

In light of the large number of publications reporting new basis sets or detailing the performance of existing sets the task of remaining informed has become very difficult for experts and nonexperts alike. The existence of such a vast multitude of basis sets is attributable, at least in part, to the difficulty of finding a single set of functions which is flexible enough to produce "good" results over a wide range of molecular geometries and is still small enough to leave the problem computationally tractible and economically within reason.

The driving force behind much of the research effort in small basis sets is the fact that the computer time required for some parts of an ab initio calculation is very strongly dependent on the number of basis functions. For example, the integral evaluation goes as the fourth power of the number of Gaussian primitives. Fortunately this is the only step which explicitly depends on the number of primitives. All subsequent steps depend on the number of contracted functions formed from the primitives. The concept of primitive and contracted functions will be discussed later.

Consider a collection of K identical atoms, each with n doubly occupied orbitals and N unoccupied (or virtual) orbitals. The SCF step increases as $(n + N)^4K^4$, while the full transformation of the integrals over the original basis functions to integrals over molecular orbitals goes as $(n + N)^5K^5$. Methods to account for correlation effects vary greatly. Only a few of the popular ones will be considered here. Second order Moller–Plesset (MP2) perturbation theory goes as $n^2N^2K^4$ but still requires an nN^4K^5 integral transformation. MP3 goes as $n^2N^4K^6$, while a Hartree–Fock singles and doubles CI will have $n^2N^2K^4$ configurations, $(n^2N^2K^4)^2$ hamiltonian matrix elements of which $n^2N^4K^6$ will be nonzero. Pople and co-workers[1] have proposed replacing the full $(n + N)^5K^5$ integral transformation with a $n^2N^4K^6$ partial transformation every CI iteration.

Ernest R. Davidson was born in Terre Haute, IN, in 1936. He received a B.Sc. degree in chemical engineering from Rose-Hulman Institute in 1958 and his Ph.D. in 1961 from Indiana University in chemistry. He was an NSF postdoctoral fellow at the University of Wisconsin in 1961. From 1962 until 1984 he was on the faculty at the University of Washington. In 1984 he returned to Indiana University where he is presently a distinguished professor and director of the Quantum Chemistry Program Exchange and of the Chemical Physics program. He was elected to the International Academy of Molecular Quantum Science in 1981. He has over 200 publications in the field of computational quantum chemistry including work on density matrices, charge, spin, and momentum distributions, electronic structure of excited states, the structure of organic radicals and diradicals, and ab initio computational methods.

David Feller was born in Dubuque, IA, in 1950. He received a B.S. degree from Loras College in 1972 and a Ph.D. in physical chemistry from Iowa State University under the direction of Professor Klaus Ruedenberg in 1979. He then joined the research groups of Professors Wes Borden and Ernest Davidson at the University of Washington as a postdoctoral associate. In 1982 he was promoted to the position of Research Associate/Faculty. Currently he holds an Associate Scientist position at Indiana University. His research interests include the application of very extended basis set ab initio techniques for determining molecular properties.

0009-2665/86/0786-0681$06.50/0

For certain ranges of n, N, and K this may represent an improvement. The MP3 step by itself only requires one iteration.

As is evident from the above discussion, even a small reduction in the size of the basis set can result in an enormous overall speedup. Since the amount of computer time in question can often be measured in days there is great interest in finding smaller basis sets that produce results equivalent to much larger ones.

Since the choice of basis set must be guided by considerations of desired accuracy in the results and computational costs, it becomes imperative that sufficient information be available so that an intelligent choice is possible. Some of the new basis sets appearing in the literature are used for a single project and then disappear into oblivion. Others are proposed, without extensive calibration, as serious contenders to replace basis sets in common use.

In practice, most popular computer programs for ab initio calculations contain internally defined basis sets from which the user must select an appropriate one. In principle this should facilitate switching from one set to another as it becomes warranted by the course of the investigation. For example, preliminary determinations of all relevant equilibrium and transition state geometries on a given potential energy surface may be made with a small basis set. Subsequent refinement at a chosen subset of these points can then be made with a larger set.

In general the evolution of "black box"-type quantum chemistry programs (such as Gaussian-82 by Pople and co-workers), coupled with the increasing performance of computer hardware, has resulted in a tremendously expanded role for ab initio techniques in academic and commercial applications. It may be possible in the future to have programs make an informed decision about the choice of basis set based on the results of thousands of previous calculations which are accessible in a data base, but that time has not yet come. The responsibility still rests with the program user.

In this article we will seek to review basis sets of the type known as "contracted Cartesian Gaussians" from the point of view of the nonspecialist who is interested in a practical guide to what is generally available today and what are some of the pitfalls to avoid. Other recent reviews on this subject from slightly different points of view have been written by Huzinaga[2] and Wilson.[3]

Boys[4] first advocated the use of Gaussian-type basis functions on the practical ground that all of the integrals required for a molecular calculation could be easily and efficiently evaluated. Because a Gaussian has the wrong behavior both near the nucleus and far from the nucleus, it was clear that many more Gaussians would be required to describe an atomic orbital than if Slater-type orbital (STO) basis functions were used. On the other hand, integrals involving STO's, at best, were expensive and, at worst, were intractible for molecules.

Fortunately, the ratio of the number of Gaussians to the number of STO's required to obtain comparable accuracy is not as large as originally believed. Although 4 Gaussians are needed to get within 1 mH (1 mH = 6.27 kcal/mol) of the exact energy for the hydrogen atom, for atoms further down the periodic table, such as argon, the ratio is more like 2.6:1. Furthermore, when "polarization" functions, e.g., d functions on carbon, are considered the ratio is reduced even more. A detailed comparison of the relative performance of STO and Gaussian d functions using a wide range of position and momentum space properties found that a single STO was worth no more than two Gaussians in the copper atom.[5] In a comparisions of STO and Gaussian basis sets in molecular property calculations Carsky and Urban[6] and Rosenberg and Shavitt[7] find similar values with similar size basis sets.

To some extent the disadvantage associated with the larger number of Gaussians compared to STO's was ameliorated through the work of Clementi[8] and Whitten[9] who introduced the concept of "contracted" Gaussians as approximations to atomic orbitals. In this scheme, the basis functions for molecules became, not individual Gaussians, but rather fixed linear combinations. Early work dealt with Gaussian lobe functions in which, for example, a p orbital was approximated by differences of s orbitals slightly displaced from each other. For higher angular momentum, the number of terms required, and the loss of accuracy from differencing in computing the integrals, made the method unwieldy.

Clementi and Davis[10] extended the use of "contracted" functions to include Cartesian Gaussians. Although individual integrals with Cartesian Gaussians are somewhat more complicated than with Gaussian lobes, the ease in extending the basis set to higher angular momentum has made this procedure the method of choice. Contrary to the misapprehensions about the use of Gaussians as basis functions, fewer contracted Gaussians than uncontracted Slater-type orbitals are required, in fact, for a given level of accuracy. One of the first widely distributed program packages to permit contracted Gaussians was IBMOL. The 1966 user's guide[11] to version 2 boasts the impressive ability to handle 700 Gaussians of s, p, d, or f symmetry on an IBM 7094 with 35,000 32-bit words of memory.

Definitions

Slater-type orbitals are defined in spherical polar coordinates as

$$\chi_{nlm} = Nr^{n-1} \exp(-\zeta r) Y_{lm}(\theta,\phi)$$

where l and m are angular momentum quantum numbers, N is a normalization constant and ζ is the orbital exponent. The main claims to superiority of these orbitals are their short- and long-range behavior. Actually, of course, a linear combination of such functions will behave correctly only if the smallest orbital exponent is exactly $(2I)^{1/2}$, where I is the ionization potential in atomic units. Thus, variationally determined orbitals will have an exponential tail with the wrong decay rate. Also, the cusp at the nucleus will only be correct for particular linear combinations of Slater orbitals.

Cartesian Slater orbitals could be defined as

$$\chi_{kmn} = Nx^k y^m z^n \exp(-\zeta r)$$

Ordinary Slater orbitals can be written as simple linear combinations of these using the definition of spherical polynomials.

Cartesian Gaussian orbitals are defined as

$$\chi_{kmn} = Nx^k y^m z^n \exp(-\zeta r^2)$$

Again, linear combinations of these can be used to form spherical polynomials of definite angular momentum and this is done in some programs. In common parlance, the whole set of Cartesian Gaussians for a fixed $L = k + m + n$ is referred to as s, p, d, etc., even though they generally contain components of lower angular momentum. The six functions with L equal 2, for example, are called "d" functions even though one linear combination of them is really an "s" function.

Unlike Slater orbitals, no finite linear combination of Gaussians can have the correct exponential tail. Similarly, no finite combination can have a cusp at the nucleus. In general, long range overlap between atoms will always be underestimated with Gaussians, as will the charge and spin density at the nucleus. While Slater orbitals may be wrong at these limits, the error is not so systematic.

Atomic orbitals are widely referred to by chemists, but are actually only loosely defined. Hydrogenic orbitals are too different from the optimum orbitals for describing an atom to furnish a suitable definition. Slater originally introduced STO's as approximations to atomic orbitals when the parameters n and ζ where chosen by certain empirical rules. Consequently, STO's are still frequently confused with atomic orbitals. In this paper we will use the term "atomic orbital" to refer to a Hartree–Fock orbital resulting from a converged (i.e., basis set limit Roothaan–Hartree–Fock or numerical SCF) calculation on some state of the atom with spin (singlet, doublet,...), symmetry (S, P,...) and equivalence restrictions (e.g., $p_x = p_y = p_z$). Unfortunately, this definition is still somewhat state dependent as the 2p Hartree–Fock orbital for the carbon s^2p^2 ^3P ground state is not identical to the 2p Hartree–Fock orbital for the s^1p^3 configuration average.

Rydberg orbitals for atoms (i.e., orbitals outside of the valence shell) can be defined by "improved virtual orbital" (IVO)[12] calculations on the ground state, or by SCF calculations on the excited state. In the latter case the Rydberg orbital would not quite be orthogonal to the valence orbitals from a ground-state calculation. Valence p orbitals for atoms such as beryllium which have no occupied p orbitals in the ground state can be defined by calculations on an excited state in which that orbital is occupied. Valence d orbitals for elements Al–Ar can be defined similarly.

Partially occupied d shells of transition metals provide an especially difficult problem. Even though s^1d^{n+1} and s^2d^n states have d orbitals with overlaps greater than 0.9, this small change in the orbital results in several eV change in energy. Thus, a single d orbital for such metals cannot usefully be defined.

Description of electron correlation in atoms is most efficiently handled by use of natural orbitals. These orbitals are the size of the occupied orbitals but have additional nodes. Such orbitals do not occur as the occupied orbitals in any excited state, so they are outside of ordinary chemical experience.

A somewhat confusing nomenclature pervades the basis set literature. What are termed "first-row" basis sets are actually intended for use on second-period elements, Li–Ne. "Second row" sets cover third-period elements and so on. In order not to introduce any additional confusion in this paper we shall continue to use this convention.

The Basic Primitive Functions

Individual Gaussian orbitals are commonly referred to as "primitive" orbitals. The first task in obtaining contracted orbitals is to choose a good set of primitives. It is generally assumed by chemists that molecules can be viewed as a collection of slightly distorted atoms. That is, the energy required to completely ionize a molecule is only slightly different from the energy required to completely ionize the constituent atoms, and the molecular charge density only differs a little from the sum of the atomic densities. Hence it is natural to require that the primitive basis set provide an accurate description of the atoms.

The first molecular calculations used a minimum set of STO's with exponents chosen by Slater's rules. Later, exponents were chosen to minimize the SCF energy of the atom. At far greater cost, the orbital exponents could also be optimized for the molecule at each point on the potential surface. Unfortunately, no calculation at this level yielded quantitative results because such a small set of Slater functions is inadequate to describe the isolated atom. Also, describing the change in size of an atom in a molecule by reoptimization of the basis functions at each geometry is prohibitively expensive.

Most Gaussian primitive sets are constructed by optimization of the Hartree–Fock energy of the atom. This choice will place heavy emphasis on representing the core orbitals, as these orbitals contribute most of the total energy of an atom. For use in a molecular calculation, this set may be supplemented with additional diffuse primitives and with functions of higher angular momentum. As these functions are designed to describe distortion of atoms in molecules, they cannot be optimized in atomic calculations.

The number of primitives used has a critical effect on the cost of certain parts of a molecular calculation. For gradient searches for optimum structures, the cost can be dominated by the time to do the basic integrals; and, hence, small primitive sets are essential. Calculations of the spectrum and other properties by perturbation theory or configuration interaction depend on the number of contracted basis functions, but are relatively independent of the number of primitive functions. Hence, in this step there is a strong motivation for using larger numbers of primitives to produce better contracted basis functions. Pople and co-workers, for example, routinely do the geometry search with a smaller basis set and then calculate the energy, at the selected geometries, with a more elaborate basis.

Calculation of other properties, such as the field gradient at the nucleus, is very sensitive to basis functions which have almost no effect on the energy. Because blanket inclusion of enough functions to describe all properties well is very expensive, special purpose functions need only be used when that particular property is desired.

As noted above, many basis sets are designed to optimize the SCF energy of the atom, and are then supplemented and/or modified for use in molecular SCF calculations. This would seem to ignore the correlation energy problem. Fortunately, the basis functions needed to describe correlation energy are localized in the same region of space as those needed for the SCF energy. Hence the same primitive set can serve almost equally well for the SCF and correlating orbitals.

The first optimized Gaussian set for atomic SCF energies was published by Huzinaga.[13] His most elaborate set had 10s and 6p basis functions and covered atomic numbers 1 to 10. The SCF error increases with atomic number and is 6 mH for neon. His (9s,5p) basis produces an error for neon of 20 mH. This basis was used by Dunning to construct his "double zeta" (DZ) basis which had an error only 2 mH larger than the error with the uncontracted primitives. The term "double zeta" derives from the early common use of the Greek letter ζ to represent the exponent of an STO basis function. A single zeta (SZ) basis was a basis in which each atomic orbital was represented by a single STO. For example carbon would have five functions ($1s,2s,2p_x,2p_y,2p_z$). Similarly a DZ basis had two STO's per atomic orbital. The nomenclature carried over into work with Gaussians, where in general usage a DZ basis means any basis set with two contracted functions per atomic orbital.

Van Duijneveldt extended Huzingaga's work through (14s,9p) basis sets. This set gave an error of only 0.2 mH for neon. No systematic contractions of the van Duijneveldt basis have been published, but Davidson[14,15] has built contractions using it for carbon and oxygen.

The effort required to develop accurate basis sets can be quite large if each component is individually optimized. For example, to produce an energy-minimized (14s,9p) set requires searching in a 23-dimensional space. The problems associated with local minima and slow convergence are such that no individually optimized exponent sets larger than van Duijneveldt's have been produced.

Inspection of the optimal exponents reveals that the ratio between successive exponents in the valence region is nearly constant. Hence, it was suggested by Ruedenberg[16,17] that an "even-tempered" or geometric sequence of exponents would represent a good approximation to the independently optimized set. If the exponents are written as $\zeta_i = \alpha\beta^i$, with different α and β values for s,p,d,... symmetry functions, the number of parameters to be optimized is greatly reduced. For example, all exponents in an (s,p) basis set on carbon can be specified in terms of only 4 parameters (α_s, β_s, α_p, and β_p) regardless of the number of primitives. The first suggestion of a geometric sequence for exponents came in work on the beryllium atom by Reeves.[18]

Furthermore, the optimized $\alpha(N)$ and $\beta(N)$ can be parameterized as functions of N, the number of primitive functions, allowing extrapolation to very large basis sets without the need for reoptimization.[19,20] This approach gives a (14s,9p) SCF energy for neon of −128.5464, 0.7 mH above the Hartree–Fock limit and 0.5 mH above the independently optimized set. The ease of generating arbitrarily large even-tempered sets has meant that some of the lowest energy SCF and CI wavefunctions for molecules such as H_2O, H_2CO, H_2S, NH_3, and HCl have been produced with this method.[21] Bardo and Ruedenberg[22] also experimented with the additional constraint that the exponents of the p-type Gaussians remain equal to some of the s exponents in a fashion similar to the basis sets of Pople.

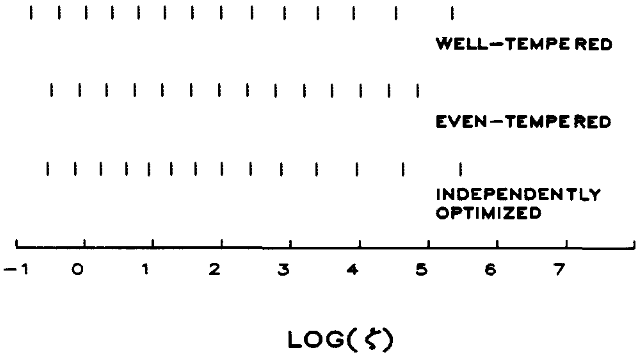

Figure 1. The distribution of s-symmetry gaussian exponents in three energy-optimized (14s,9p) neon atom basis sets.

Huzinaga and co-workers[23] have recently devised a four-parameter function of the form:

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

for generating Gaussian exponents. They call their basis "well-tempered" after the parent even-tempered formula. 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 of −128.5467, 0.4 mH above the Hartree–Fock limit and only 0.1 mH above the independently optimized set. In Figure 1 the relationship between independently optimized, even-tempered, and well-tempered exponents, ζ_i, is shown for the s-symmetry functions in neon. The original paper reported (14s,9p) sets for the first-row elements and (16s,11p) sets for the second row. Subsequent papers dealt with the elements K through Xe[24] and special considerations of the first- and second-row transition metals.[25] 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.

Silver, Nieuwpoort, and Wilson[26,27] have carried this logic one step further and argued that universal, even-tempered basis sets can be devised which apply equally well to all atoms. Of course, this requires very large basis sets if a large range of atomic charges are to be covered with equal accuracy.

An interesting point in this regard is the question of completeness. It is desirable for high accuracy studies to be able to define a sequence of basis functions for which the limit of the sequence is a complete set. It was noted that the even-tempered primitive set for a fixed value of β when extended to an infinite set is incomplete. For example, if the optimized hydrogen (4s) set, which yields an energy of −0.4987 H, is extended with both larger and smaller exponents (using the same β) the energy converges to −0.4995 H rather than −0.5. However, it can be shown[28] that when β is varied with N, as in the formula of Schmidt and Ruedenberg, the infinite set is complete. Actually, the basis set will become over-complete (i.e. linearly dependent) in this limit. For practical purposes β must not be too close to unity or the calculation becomes numerically unstable due to near linear dependency.

Clementi and Corongiu[29] combined the ideas of even-tempered exponents and universal basis sets along with six constraints to produce a new type of basis set, which they called "geometrical", for use in large molecule calculations. All atoms from H to Sr are represented by the same set of exponents (although differing numbers of s, p, and d functions appear). Three sets were reported, all with $\beta = 3.0$ and values of α in the 0.012–0.016 range. Examples of variation in the composition of the three sets for selected elements are: 6s–10s on hydrogen, (9s,5p)–(13s,8p) on carbon and (15s,8p,5d)–(19s,11p,8d) on iron. 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.

Another approach to selection of the primitive basis is the STO-*n*G basis. This was formed by a least-squares fit of primitive Gaussians to a minimum set of Slater type orbitals. The orbital exponents for the Slater orbitals were then chosen as the average best exponents from molecular calculations. Such basis functions produce rather poor atomic energies.

The last approach which will be mentioned here is the simultaneous optimization of the exponents and contraction coefficients based on atomic SCF energies. Whitten's original Gaussian lobe scheme, for example, optimized a (10s,5p) primitive set for a contraction to 3s and 1p basis functions. Pople's various basis sets of the *n*–*ij*G type (e.g., 6-31G) are of this type with the added constraint of shared orbital exponents between primitives on the same atomic shell. For example, the *i*'th exponent carbon 2s exponent would be identical with the *i*'th 2p exponent. Some of these basis sets are commonly used as optimized, while others are normally rescaled for use in molecules. No basis of this type is reported for neon, but the SCF error for fluorine is 48 mH with the 6-31G basis.

The Choice of Additional Primitive Functions

For accurate molecular calculations all of these atomic basis sets require supplementation. The most obvious defect is that atomic calculations only define functions with the same L as the occupied orbitals, e.g., s and p for carbon. Polarization functions, which describe the distortion of the atomic orbitals in a molecular environment, are often determined by choosing an average value from a group of energy-optimized functions computed for a small, representative set of molecules. The energy loss in choosing an average value, as opposed to explicitly optimizing for each new molecule, is small due to the generally observed insensitivity of the energy to the polarization function exponent.

When sufficiently many polarization functions are employed exponents optimized for atomic correlation purposes will essentially span the same space as those needed for polarization and the correlation set can serve dual purposes. In the case of carbon, for example, the energy-optimized exponent for a single d-type Gaussian to be used for angular correlation purposes in the atom has a value of roughly 0.6 compared to average values in the 0.7–0.8 range taken from molecular SCF optimizations where the function is used for polarization purposes. Similar results for other first-row atoms suggest that the energy-optimal exponents for polarization purposes are sufficiently close to the optimal correlation exponents that the atom-determined exponents may be transferred to the molecular case with no significant loss in accuracy. In a study of electron affinities and first-ionization potentials Feller and Davidson[30] have reported exponents for multiple d, f, and g sets on carbon, nitrogen, and oxygen determined at the singles and doubles CI level.

From a practical point of view the convergence of the total energy and many other properties is quite slow. An extensive study by Ahlrichs and co-workers[31] of the impact of higher polarization functions on F_2 showed that the relative importance of various types (d,f,g...) of polarization functions varies with the computational method chosen. While SCF and multiconfiguration SCF (MCSCF) methods were energy converged at the (2d,1f) level, multireference singles and doubles CI (MR SD–CI) required up through g functions. The largest calculation, a 280,000 configuration MR SD-CI using an (s,p,d,f) basis, obtained $D_e = 1.59$ eV (expt = 1.66 eV) and $R_e = 1.420$ Å (expt = 1.412 Å). Another study[32] on Cl_2 found even slower convergence with respect to higher L functions for this second row diatomic than for F_2. A single g set contributes 0.15 eV to the dissociation energy. Similar studies of the effects of higher polarization functions on the one-electron properties of oxygen and sulfur hydrides[33] have been performed. In general, higher angular momentum functions are found to be much more important for correlation purposes than for polarization effects. The addition of such functions to one atom in a molecule usually necessitates their addition to all atoms, for the sake of balance. Thus, the decision to include polarization functions can greatly increase the size of the basis.

Lists of polarization exponents have been given by many authors. One of the most extensive, due to Roos and Siegbahn,[34] lists values for the entire first and second rows. Recommended exponents for multiple d-type polarization functions and the relative merits of f-type functions in second-row atoms have recently been discussed by Magnusson and Schaefer.[35] Roos and Sadlej[36] have proposed a method for determining polarization function exponents, called electric-field variant Gaussians, which are claimed to provide a more balanced treatment of molecular properties other than the total energy. Early results on LiH are nearly identical with properties obtained from numerical HF calculations. However, no list of exponents for other atoms was given.

A somewhat less popular method for introducing polarization effects into molecular wave functions is to add lower angular momentum functions at the bond centers.[37] 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[38,39,40,41] have suggested exponents for these functions and have advocated their use in SCF and correlated wave function calculations.

If a wide range of the molecular potential energy surface is of concern the positioning of the bond functions can be a problem. Bauschlicher[42] has warned of another problem associated with bond functions. In an effort to understand why bond functions seemed to perform better than atom-centered polarization functions at predicting diatomic dissociation energies he studied the ClO molecule with both types of basis sets. The bond-centered sets were found to have CI superposition errors as large as 0.7 eV, compared to standard polarized set errors of 0.09 eV. That is, the difference in computing the atomic energy with just the atom-centered functions as opposed to the atom-centered plus bond functions was 0.7 eV. Thus, a large part of the computed molecular dissociation energy was due to the lowering of the atomic energy from the nearby additional functions.

Wright and co-workers[43,44] have argued in favor of exploiting the superposition error by balancing it against the basis set incompleteness error in an empirical manner so as to allow the calculation of dissociation energies in molecules which are too large to be handled otherwise. However, when correlated wave functions are being used, as in SCF + CI calculations, Bauschlicher[45] cautions that this balance is very difficult to maintain as the geometry and correlation recovery method are varied.

Less obvious than the lack of polarization functions in an atomic basis set is the fact that the valence shell of energy-optimized basis sets is usually relatively inflexible as most of the basis functions are used in getting the core energy correct. To correct this deficiency it is often desirable to add a more diffuse basis function to all optimized sets.

The Choice of Contraction Coefficients

As in the case of primitive functions, the contraction schemes represent a compromise between accuracy and feasibility. Without loss in accuracy, all basis functions which appear in the same ratio in all molecular orbitals of importance can be grouped into a fixed contraction. If the core electrons are not to be correlated in the final wave function, and if no core hole states are being considered, then it is usually safe to contract the core orbitals. Tight basis functions for the core also appear in the valence orbitals because of orthogonality, but experience has shown that they have the same coefficient ratios in the valence orbital as they have in the core orbitals. The cost per integral increases rapidly with the number of terms in the contraction, so for integral-limited computing steps such as SCF gradient geometry searches there is a strong motivation to use rather crude approximations to the core orbitals. On the other hand, large CI studies of vertical excitation energies are limited by the number of basis functions, rather than the number of primitives, so it is better to use an accurate description of the core.

Almost all contractions are based on atomic Hartree–Fock calculations. Occasionally, contracted polarization functions are formed from calculations on molecules, but these do not transfer accurately to other molecules. Contractions may be segmented, that is the primitives are partitioned into disjoint or almost disjoint sets. Alternatively, Raffenetti[46] has proposed "general" contractions in which every primitive may appear in every contraction.

In Raffenetti's scheme the suggested minimal basis set would consist of nothing more than the atomic SCF orbitals for each atom in the molecule. These would be supplemented by making the more diffuse primitives uncontracted. Feller and Ruedenberg have shown for a variety of molecules that the SCF AO's overlapped the space of the "optimal"[47,48] minimal basis set to better than 0.99.

For example, a carbon [4s,2p] contraction suitable for molecular calculations would consist of the 1s, 2s, and 2p atomic orbitals from an uncontracted, isolated-atom calculation done with energy optimal exponents, plus the two most diffuse s-type and single most diffuse p-type primitives. Slight complications can arise in uncontracting the "most diffuse" primitives if the even-tempered set has been extended at the diffuse end for the purpose of improving certain properties. In that case it would be advisable to keep the extra diffuse Gaussian primitives uncontracted along with the diffuse primitive or primitives from the energy-optimized set even though the computational cost has been increased.

Without loss in energy the uncontracted primitives could be dropped from the 1s, 2s, and 2p orbitals, thus reducing somewhat the cost of the general contraction for the remainder of the 1s and 2s functions. In fact, for most small to medium size energy-optimized basis sets one can span almost the same function space by simply uncontracting the three most diffuse s-type primitives because all but about three of the primitives are being used to describe the core orbital. In general, the closer the basis set is to being a minimal basis the greater is the advantage to a general contraction approach in terms of the ease of generating the contraction coefficients for any atom. Whereas, as one proceeds to more extended basis sets with long inner shell expansions and many uncontracted Gaussians, the less it offers and the more its computational expense increases. The general contraction scheme is also quite expensive to use in geometry searches, but it may offer advantages in terms of interpreting the wave function. The use of generally contracted basis functions has increased slowly because many integral programs do not handle such functions efficiently.[49]

Two approaches have been suggested for calibrating basis sets. The first, exemplified in the work of Pople and co-workers, uses the same level of theory for many molecules and assumes that the average error is representative of the actual error for similar molecules. Pople has extensively tested certain contractions so, for these contractions, some idea of the average error to be expected in certain applications is available. Unfortunately, Pople has focused on a rather narrow set of properties, such as geometry and heats of isodesmic reactions, and his earlier small basis sets may not do well for other properties such as the dipole moment. Also, his earlier work compared only SCF results with experiment. Often the error in the geometry seemed to increase as the basis set was improved toward the SCF limit.

The other approach is to generate a sequence of basis sets which rapidly approach an effectively complete set for the property of interest, and then to do a sequence of calculations with these sets, extrapolating to the limit of a complete set. A prerequisite for effective extrapolation is the smooth behavior of the property as a function of the basis set size. Figure 2 demonstrates that not every sequence of basis sets produces acceptable results. The dipole moment of carbon monoxide was computed at the SCF level with a variety of uncontracted even-tempered basis sets to which a single diffuse s and p function had been added. The same property was computed with the STO-3G, *n*-*i*1G and *n*-*i*11G (s,p) basis sets. It is evident that the two sequences are approaching the same limit, but in the small basis set region the results can be erratic. Since

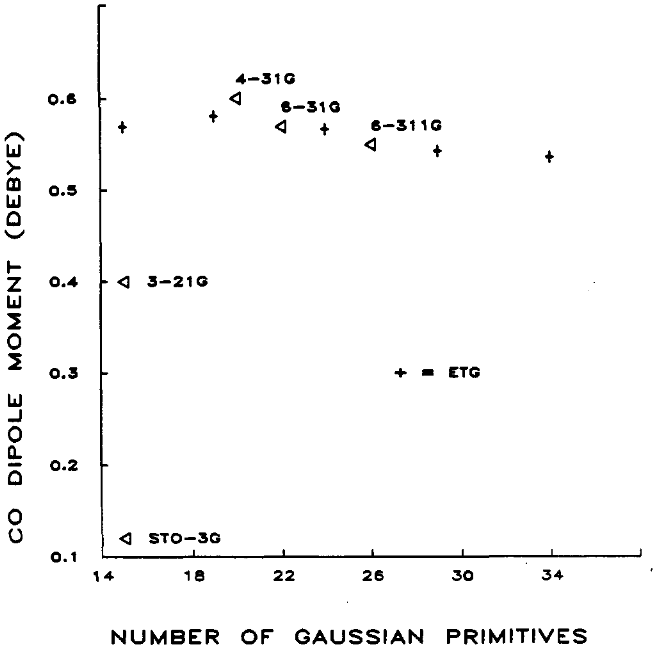

Figure 2. Variation in the SCF dipole moment of carbon monoxide as a function of the number of gaussian primitives for various basis sets.

the experimental CO dipole moment is −0.112 D (C^-O^+), the STO-3G basis value might be considered "best". Comparison of small basis set SCF results with experiment can often lead to this mistaken conclusion.

Because no finite basis set can be complete in all regions of space, it is necessary to select the sequence carefully for the property being considered. Of course, the basis set must be good enough that refinement of the basis in one region does not produce distortions in the overall charge distribution. In the context of ab initio calculations on heteronuclear diatomics Mulliken[50] warned of the need for "balanced" basis sets. Although it is straightforward to insure a balanced set when all atoms in a molecule are from the same row of the periodic table, when atoms from two or three different rows are involved and the basis set is small, the task increases in complexity.

The question of how to measure the overall quality of a basis set is a difficult one. Wilson[51] has recently proposed the use of the orbital amplitude difference function as a means of probing basis set deficiencies. For two orbitals, one expanded in M basis functions, the other in N basis functions, the difference function is given simply as the difference of the two orbitals, $\phi^M(r) - \phi^N(r)$. Maroulis et al.[52] have made use of information theoretical techniques in an attempt to classify and improve basis set performance with regard to properties other than just the energy. Their study included the total energy, the bond length, and four electric moment tensors in some simple diatomics and water.[53] Averaging over several vibrational and rotational states was considered for the electric moments.

The magnitude of the reduced gradient of the energy with respect to the gaussian exponents,

$$|g| = \sqrt{\Sigma G_i/n}, \text{ where } G_i = \partial E/\partial \ln \zeta_i,\ n = \text{no. of exponents},$$

has been suggested as a measure of basis set quality by Kari, Mezey, and Csizmadia.[54] For any basis set this norm will go to zero provided the exponents have been perfectly optimized. Some properties, such as charge density at the nucleus, appear to be fairly sensitive to the degree of convergence in the exponents. Based on this idea four groups of "uniformly balanced" basis sets, from (4s,2p) up to (10s,5p), were prepared for the first-row atoms.[55] However, while $|g|$ permits basis sets of equal composition to be compared, it is inherently incapable of being used to judge sets of different size in order to determine a practical measure of their convergence to the complete basis set limit. Perhaps a better measure set quality would be the maximum eigenvalue of the second derivative matrix, $(\partial^2E/\partial \ln \zeta_i \partial \ln \zeta_j)|_{\text{opt}}$ which provides a feel for how sensitive the energy is to the given choice of exponents.

Nonetheless, the great diversity of bonding situations which may arise for arbitrary molecular geometries coupled with the ever-expanding set of experimentally accessible properties makes it very difficult to predict in advance the level of accuracy in an ab initio calculation. In the absence of a generally agreed upon measure of the "quality" of a basis set researchers must still rely on their ability to test a proposed basis set on a prototypical molecular system for which the property in question is already known.

The Basis Set Superposition Error

If the property to be computed is the dissociation energy, then particular care must be paid to the "basis set superposition error" (BSSE). The severity of this problem is proportional to the incompleteness of the atomic basis set. Thus, to the extent the atom-centered basis is inadequate, the SCF or CI procedure will attempt to utilize any available basis functions on neighboring centers to make up for the deficiency. On the other hand, many molecular properties do not appear to be particularly sensitive to the BSSE.[56]

In the simple case of a diatomic an artificial lowering of the two atomic energies in the vicinity of the equilibrium geometry will masquerade itself as true binding energy and skew the calculated well depth. The BSSE will be especially severe when moderately diffuse basis functions are included in the valence set and crude approximations for the core orbitals have been used.

No completely reliable scheme for either eliminating or estimating superposition errors has been given. Perhaps the most popular prescription for partially resolving the problem is the counterpoise method of Boys and Bernardi.[57] This procedure corrects the atomic energies by computing the atoms in the full molecular basis set. As pointed out by several authors[58,59] this will overestimate the BSSE because it will provide the atom with basis functions that are "used up" by the other atoms in the molecule.

Due to the shallowness of the potential curves (often on the order of 1 kcal/mol or less) studies of van der Waals complexes are particularly subject to this problem. The fact that correlated wave functions are necessary to find any bonding at all merely compounds the problem. An illustration of the inherent difficulties caused by the BSSE in van der Waal's complexes is provided in a recent paper on Mg_2[60] which is experimentally known to have a well depth of 1.229 ± 0.003 kcal/mol and an R_e = 3.89 Å. Basis sets as large as (14s,11p,4d,2f) contracted to [9s,5p,2d,1f] were used in conjunction with various kinds of many body perturbation theory (MBPT) up to fourth order. Even at this level the authors conclude that the accuracy of the results "depends considerably on the magnitude of the basis set superposition error."

An SCF study[61,62] of the interaction energy of HF + HF with 34 different basis sets, ranging from minimal to very extended, explored the effectiveness of the full counterpoise correction and a limited approach in which only the polarization functions on neighboring centers was used to augment the atomic basis. The authors concluded that the counterpoise-corrected interaction energy was no more reliably accurate than the uncorrected energy. They furthermore conclude that other schemes for estimating the BSSE which include only the unoccupied orbitals from neighboring centers do not in general yield accurate corrections. At the correlated level on a much simpler system (He + He) Gutowski et al.[63] find that the full counterpoise method produced better results than methods which try to omit the occupied orbitals from neighboring atoms.

Basis Sets Used with Effective Core Potentials

The tremendous cost of ab initio calculations has motivated many attempts to find computational shortcuts. One such approach is based on the observation that core orbitals are relatively inert to changes in chemical bonding (the so called "frozen core approximation"). Another observation is that the effect of core electrons on the valence electrons can be treated through the use of a potential energy term expressed as the sum of local functions multiplied by projection operators. Based on these two assumptions effective core potentials (ECP's) or pseudopotentials, as they are sometimes called, reduce the computational problem to dealing with valence electrons only. Thus, for example, germanium becomes as easy to treat as carbon.

Most early results obtained from ECP's compared favorably with results obtained from all-electron calculations, although there was a tendency to find shorter bond lengths and somewhat deeper potential energy curves. Recently developed ECP's[64] have reportedly solved these problems so that ECP and all-electron results are now in nearly exact agreement.

In principle it should be possible to use the valence portion of an ordinary all-electron basis set with an ECP, but the results are somewhat uneven. Most authors derive special valence basis sets by least-squares fitting of nodeless pseudo-orbitals that match the usual atomic valence orbitals for all values of r greater than some chosen critical value. This raises the question of whether extended basis sets can be used with ECP's.

ECP's and companion basis sets featuring shared s and p exponents for first- and second-row atoms have been given by Stevens et al.[65] and by Topiol et al.[66] Nonrelativistic ECP's for the transition metals Sc to Zn and relativistic ECP's for Y to Hg[67] and main group elements Na to Bi[68] have been constructed by Hay and Wadt who point out that even for first row transition metals the relativistic correction to the $4s^23d^{n-1} \rightarrow 4s^13d^n$ excitation energy can be as large as 0.4 eV. For third-row transition elements the correction increases to 3.2 eV. In order to account for the dominant relativistic effects they incorporated the mass-velocity and one-electron Darwin corrections into their potentials.

In some circumstances the division of electrons into "core" and "valence" sets is not so clear cut. For example, while the 3s and 3p orbitals are usually treated as part of the core for first-row transition metals, they are roughly the same size as the 3d valence orbitals and in some cases may need to be treated along with the valence space. In order to accommodate such situations Hay and Wadt[69] have developed ECP's which do not include the outermost core orbitals of their previous potentials.

Aside from not having to contend with the core electrons the valence basis set requirements for ECP and all-electron calculations are the same. An interesting example of the carry over from all-electron work to work with ECP's appears in an article[70] on SeO_2 and TeO_2 done with the Hay and Wadt potentials. The importance of d-type polarization functions had been demonstrated almost 16 years ago for SO_2,[71] the third-period analogue of these dioxides. For SeO_2 and TeO_2 the presence of d-polarization functions was found to be critical to achieving bound systems.

Other collections of ECP's have been published by Bachelet et al.[72] including almost the entire periodic table, and Rappe et al.[73] for elements Mg to Cl.

Selected Contracted Basis Sets for All-Electron Calculations

In this section we shall briefly describe a few of the basis sets enjoying current widespread use. For those interested in experimenting with basis sets on their own, the best single source compilation of Gaussian exponents and contraction coefficients is probably the recent book by Poirier, Kari, and Csizmadia.[74]

STO-3G. Perhaps the most widely used basis set of all time is the STO-3G minimal basis developed by Pople and co-workers. Originally formulated for first-row elements[75] it was later extended to second-row,[76] third-row,[77] and fourth-row[78] main group elements. It has also been applied to first- and second-row transition metals.[79] The main attraction of this basis, other than its small size, is its effectiveness in predicting geometries. Pople[80] reports that the mean absolute deviation from experiment for SCF bond lengths in several dozen molecules containing H, C, N, O, and F is 0.030 Å. Moreover, the literature contains hundreds of other comparisons with experiment, most of which fall in a similar range.

Such agreement may appear remarkable for a minimal basis, until it is realized that the large STO-3G basis set superposition error helps cancel other defects to produce reasonable bond lengths. For example, when the SCF bond length of CO is determined with a basis set of accurate Hartree–Fock (1s,2s,2p) atomic orbitals the result is 0.11 Å longer than experiment, while the STO-3G determined SCF value falls within 0.02 Å. In the case of F_2 very large, polarized basis sets find an SCF bond length of 2.50 Å (expt = 2.68 Å) and an energy of atomization of −1.37 eV (expt = +1.68 eV), while the STO-3G basis gives 2.49 Å and +0.20 eV. That is, due to fortuitous cancellation of error, the results are better than the Hartree–Fock model is capable of yielding. In general, the smaller the basis the more ab initio calculations assume an *empirical* flavor.

As might be expected when agreement with experiment relies heavily on cancellation of error the agreement will be much worse in some cases. Transition metal bonds are such a case, with deviations of 0.2 Å or larger being reported.

Energetics are not predicted with the same accuracy as geometric parameters. The restriction to a single factor for both σ and π p functions results in π bonds which are too weak relative to σ bonds. For example, the reaction $C_2H_6 + C_2H_6 \rightarrow 2(CH_4) + C_2H_4$ has an equal number of CH bonds in the reactants and products but two C–C single bonds on the left and one C=C double bond on the right. The STO-3G basis overestimates the SCF ΔE for this reaction by 33 Kcal/mol compared with large, polarized basis sets.

4-21G and 3-21G. These "split valence" basis sets were a compromise between the speed obtainable using the STO-3G basis and the accuracy of larger, slower sets. The 4-21G basis[81] was designed for efficient use in geometry optimizations which utilize analytical gradients. The 3-21G basis sets have been constructed for first-[82] and second-row[83] elements of the periodic table. Performance characteristics for the 4-21G and 3-21G sets are very similar. The reported mean absolute deviation from experimental bond lengths is 0.016 Å for first-row compounds and 0.071 Å for second-row compounds with the 3-21G basis. Skewing of σ bond energies is significantly reduced compared with the STO-3G basis, so that the SCF ΔE for the isodesmic reaction described above is in essentially exact agreement with the result from larger basis sets.

Two further extensions of the 3-21G basis have been published. The first is a set of polarization functions for the second-row elements to make the 3-21G(*) basis.[84] Compared with the original set the mean absolute deviation in bond lengths dropped from 0.071 Å to 0.027 Å. The 3-21+G and 3-21++G basis sets[85] add diffuse functions to improve the description of anion geometries.

4-31G, 5-31G, and 6-31G. By increasing the number of primitives devoted to the core and first valence function, the 4-31G,[86] 5-31G, and 6-31G[87] bases improve upon 3-21G energetics at the expense of increased computer times. Adding all six components of a cartesian d function for first-row atoms gives the 6-31G* basis. Further addition of p functions to hydrogen results in the 6-31G** basis. Hehre and Latham[88] extended the 4-31G basis to second-row elements. This set is sometimes referred to as the 44-31G basis. Likewise, the 6-31G* basis is available for the first and second rows.[89] A slightly smaller set, called the 6-31+G set by Radom and co-workers[90] replaces the six-component cartesian d's with their five-component spherical counterparts. Agreement with experiment for the 6-31G* bond lengths using 33 sample closed-shell molecules was 0.033 Å at the Hartree–Fock level and 0.018 Å at the MP2 level of theory.[91] This set of molecules included some notoriously difficult cases where agreement was much worse than normal. If these are ignored, the mean absolute deviation falls to 0.013 Å at the HF level.

Augmenting the 4-31G basis with a set of diffuse s and p functions, to give a 4-31+G set, has been suggested for improving the description of anion proton affinities.[92] A comparison of calculated and experimental proton affinities for more than a dozen anions shows improvements of 10–40 kcal/mol compared to the 4-31G basis. However, this set does not provide improved anion geometries.

6-311G.** All previously described Pople basis sets have been optimized at the Hartree–Fock level. These sets are somewhat deficient for use in post HF calculations, so a 6-311G** basis[93] (single zeta core, triple zeta valence, and polarization functions on all atoms) was developed which performs better at the MP2 level. Geometrical parameters obtained with this basis at this level were found to agree almost exactly with geometries obtained with an uncontracted (8s,4p,1d/4s,1p) basis. Agreement with experiment at the MP4 level was in the 0.006–0.013 Å range, and using SD–CI wave functions[94] were generally better than 0.007 Å.

More Extended Pople Basis Sets. If accuracy beyond that obtainable with the 6-311G** basis is desired, more flexibility must be introduced into the polarization space. The set must also be supplemented with diffuse functions for description of anions. With these goals in mind Pople and co-workers[95] have developed several basis sets, the largest of which goes by the unwieldy name of 6-311++G(3df,3pd). The "++" indicates that there are additional diffuse functions on all atoms. The "3df" means that three 5-component d functions and one 7-component f function are included on all first-row atoms. Exponents for the f functions were optimized at the UMP4 level of theory. The "3pd" means three p functions and one d function on hydrogen. UMP4 ΔE's for several simple reactions improved, with respect to the 6-31G** results, by as much as 14 kcal/mol with inclusion of all the extra functions.

Dunning Double Zeta with Polarization. One of the earliest high-quality collections of contracted basis functions was due to Dunning.[96] When coupled with the suggested d polarization functions, this [4s,2p] segmented contraction of the Huzinaga (9s,5p) exponents for first-row atoms provides ample flexibility to describe most bonding situations at a semiquantitative level.

Dunning/Hay Split Valence and Double Zeta. By decreasing the flexibility of the core description in the DZ basis sets Dunning and Hay[97] fashioned a [3s,2p] or "split-valence" contraction which has been widely used. No systematic examination of the errors for this set has been collected, but errors in geometric parameters and properties would be expected to be comparable to those found with the 6-31G basis. The SV basis sets covered elements H–Ne. The same paper also contained [6s,4p] double zeta contractions of the Huzinaga (11s,7p) primitives for elements Al–Cl.

MINI-*i*, MIDI-*i* and MAXI-*i*. Tatewaki and Huzinaga have published a large collection of minimal and split valence basis sets for first-row atoms which do not restrict the 2s and 2p exponents to be equivalent. The contraction coefficients were determined via the SCF procedure. The MINI-*i* (i = 1–4) have varying numbers of Gaussians in the 1s, 2s, and 2p contracted functions. For example, MINI-1 could be denoted (3s,3s/3p), while MINI-4 is a (4s,3s/4p) contraction. By splitting the valence space functions you obtain the MIDI-*i* sets, where, for example, MIDI-1 is a [3s,2s,1s/2p,1p] contraction.

The common demominator of the MINI and MIDI contractions is the use of 3 Gaussians to represent the 2s space. In the MAXI-*i* sets 4 Gaussians are used along with up to 7 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.

McLean/Chandler Second-Row Sets. McLean and Chandler[98] have developed minimal, double, and triple zeta (for the 3p, double zeta for the rest) contracted basis sets for neutral Na–Ar and the P^-, S^-, and Cl^- anions. Huzinaga's (12s,8p) and (12s,9p) exponent sets served as a starting point. Although the authors advocate the use of general contractions as being "clearly superior", these contractions are of the segmented kind (minimal duplication of the primitives in the contracted functions) because "the bulk of current molecular computation is still being done using integral generators which limit the number of primitives in a contracted basis function".

Transition-Metal Basis Sets. Until quite recently the number of high-quality basis sets for the transition metals, Sc–Zn, was very limited, as contrasted to the abundance of sets for the first and second row elements. This disparity can be traced to unique problems connected with the transition metals as well as the simple increase in computational effort required to deal with much larger sets of primitives. A relatively large (15s,8p,5d) set has been reported by Basch et al.[99] Wachters[100] (14s,9p,5d) set for third-row elements up to Zn was developed as an extension of earlier first- and second-row basis sets. With similar contraction lengths the Wachters set produces much lower energies than the set by Basch et al. For example the [8s,4p,2d] Mn (^6S) SCF energy with the Wachters set is −1149.7872 compared to −1149.5642 with the latter. Both the Wachters' set and a smaller (12s,6p,4d) set, developed by Roos et al.,[101] have been augmented by Hay[102] to increase the flexibility in the d space. Hay has shown that the 3d orbitals of the $4s^23d^{n-2}$, $4s^13d^{n-1}$ and $3d^n$ electronic configurations vary enough that use of basis sets optimized for one can lead to errors of several eV in computed excitation energies relative to the Hartree–Fock limit. This behavior has no counterpart 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.

Just as d functions are known to be important for first and second row elements, f-type polarization functions have been shown to be critical for describing bonding between transition elements. McLean and Liu[103] found that such functions contribute over 1 eV to the binding energy of Cr_2. On the other hand, to the extent that the bonding is more 4s in nature the addition of f functions should hardly affect the binding energy. Walsh and Bauschlicher[104] evaluated the performance of various f contractions for Cr_2 and the Cu atom at the SCF and CI levels. While a single f function did well at the SCF level, when correlation is included, a two-term contraction was found to provide much better results.

Supplemental functions for Huzinaga's[105] second-row transition element basis sets have been given by Walch et al.[106] Three areas of weakness in the original sets were indicated (1) the 5s → 5p near degeracy necessitated additional 5p functions; (2) a balanced description of the $5s^24d^n$, $5s^14d^{n+1}$, $4d^{n+2}$ states require diffuse 4d functions; and (3) a set of 4f functions is needed to correlate the 4d orbitals. The Huzinaga sets ranged in size from (14s,9p,5d) to (17s,11p,7d).

Miscellaneous First-Row Sets. As an alternative to the STO-3G minimal set, whose exponents and contraction coefficients were determined by least-squares fitting, Ditchfield et al.[107] published a collection of minimal basis sets for first-row elements whose parameters were chosen so as to minimize the total energy. The contraction coefficients were determined by nonlinear minimization. One set imposed the restriction that the 2s exponents equalled the 2p exponents. Tavouktsoglou and Huzinaga[108] have developed a closely related set in which the contraction coefficients were chosen by the SCF procedure. A somewhat refined version of this method was used to generate basis sets for the transition metals Sc–Zn.[109]

Miscellaneous Second-Row Sets. Veillard[110] has published energy optimized (12s,6p) and (12s,9p) sets for Na–Ar. Orbital expansion coefficients are given along with the atomic energies resulting from various ways of contracting the primitives. Care should be exercised in using the Na and Mg sets, since they lack sufficient diffuse p functions for most molecular work. This is a general phenomenon caused by the lack of an occupied 3p orbital in the atomic ground state.

Roos and Siegbahn[111] have published a small (10s,6p) energy optimized collection of basis sets for second-row elements. This same article contained (7s,3p) sets for first-row elements.

Miscellaneous Third-Row Sets. Recently, Tatewaki and Huzinaga[112] reported two new minimal basis sets, called STD-SET(1) and DZC-SET(1), for the first-row transition elements which yield atomic orbital energies as good as, or better than, DZ basis sets. In a subsequent paper[113] they discuss ways to split the valence space so as to increase the basis sets' flexibility.

Miscellaneous Fourth-Row Sets. Stromberg et al.[114] have published (15s,11p,6d) exponent sets for the fourth-row main group elements, In–Xe. Atomic orbital coefficients are listed so that it is possible to construct DZ contractions.

An Illustrative Example

A selected subset of the basis sets described above were chosen for use in calculating the properties of formaldehyde so that some direct comparison among basis sets could be made. Recognizing that no single molecule or set of properties can illustrate all of the strengths or weaknesses of any basis, a compilation of the results from various basis sets for a well-behaved, ground-state singlet molecule was still felt to be informative.

All formaldehyde calculations were performed at the experimental geometry of Oka and Takagi[115,116] (R_{CO} = 1.20785 Å, R_{CH} = 1.1160 Å, HCH angle = 116.52°).

The large basis set limit was approximated by a (19s,10p,2d,1f) primitive set contracted to [10s,5p,2d,1f] on oxygen and carbon using energy-optimized even-tempered (18s,9p) sets extended by one additional s and p function at the diffuse end. The hydrogen basis was a (10s,2p,1d) contracted to [4s,2p,1d]. Contraction coefficients were taken from the 1s and 2p atomic orbitals on the respective atoms. s-symmetry components

TABLE I. Selected Energy-related Properties[a] of Formaldehyde ($^1A_1$) at the Experimental Geometry

				orbital energies		vertical ionization		verticle excitations		
basis[b]	total energy	energy of atomization	BSSE[c]	$\epsilon(1b_1)$	$\epsilon(2b_2)$	$^2B_1$	$^2B_2$	$^3A_2$	$^1A_2$	$^3A_1$
Dunning/Hay	−113.8292	0.349	0.0010	−0.538	−0.443	0.465	0.355	0.089	0.103	0.156
SV	−114.0542	0.481				0.532	0.381	0.131	0.150	0.223
Dunning/Hay	−113.8939	0.413	0.0011	−0.534	−0.440	0.458	0.354	0.103	0.116	0.167
SVP	−114.2255	0.555				0.526	0.405	0.133	0.150	0.224
Dunning DZP	−113.8940	0.413	0.0011	−0.534	−0.440	0.457	0.354	0.103	0.116	0.167
	−114.2253	0.554				0.525	0.383	0.133	0.149	0.224
STO-3G	−112.3537	0.418	0.0533	−0.447	−0.352	0.404	0.275	0.078	0.097	0.155
	−122.4988	0.543				0.460	0.288	0.138	0.153	0.229
3-21G	−113.2204	0.400	0.0114	−0.527	−0.430	0.462	0.349	0.092	0.108	0.162
	−113.4573	0.498				0.526	0.374	0.138	0.152	0.229
6-31G	−113.8069	0.355	0.0024	−0.534	−0.440	0.464	0.353	0.090	0.104	0.157
	−114.0408	0.478				0.527	0.378	0.131	0.144	0.222
6-31G**	−113.8680	0.415	0.0024	−0.531	−0.436	0.456	0.352	0.103	0.117	0.167
	−114.2031	0.570				0.523	0.381	0.135	0.159	0.250
MIDI-4	−113.6848	0.349	0.0018	−0.534	−0.441	0.463	0.354	0.091	0.105	0.159
	−113.9177	0.484				0.527	0.375	0.143	0.144	0.222
HF AO	−113.6604	0.163	0.0002	−0.596	−0.498	0.550	0.428	0.072	0.092	0.155
minimal	−113.7873	0.273				0.596	0.438	0.119	0.136	0.219
Atomic NO	−113.9034	0.404	0.0002	−0.534	−0.441	0.460	0.357	0.104	0.118	0.169
SVP	−114.2454	0.533				0.524	0.384	0.129	0.156	0.220
(s,p,d,f)	−113.9202	0.420	0.0002	−0.537	−0.443	0.457	0.356	0.102	0.115	0.167
extended	−114.3317	0.569				0.532	0.397	0.139	0.153	0.233
Experimental	−114.562[d]	0.60				0.53	0.40	0.12	0.15	0.22

[a] Energies are expressed in hartrees. For each basis set entry SCF values are given in the first row, followed by the frozen core CI values in the second row. The calculations were performed at the following geometry: C (0.0, 0.0, −1.141998), O (0.0, 0.0, 1.140502), H (0.0, ±1.793554, −2.251522), with coordinates given in bohrs. [b] The total number of functions associated with each basis is as follows:

Dunning/Hay SV (22)
Dunning/Hay SVP (40)
Dunning DZP (42)
STO-3G (12)
3-21G (22)
6-31G (22)
6-31G** (40)
MIDI-4 (22)
HF AO's (12)
atomic NO's ETG (40)
Extended ETG (120)

The s components of the cartesian d functions were deleted in the extended ETG basis. [c] Defined as the difference of the spin- and symmetry-restricted atomic SCF energies computed in the atomic basis and the symmetry-unrestricted SCF energies in the full molecular basis. The electronic configurations for carbon and oxygen corresponded to $1s^22s^22p_y{}^12p_z{}^1$ and $1s^22s^22p_x{}^22p_y{}^12p_z{}^1$, respectively ($C_{2v}$ symmetry imposed), with the x direction pointing out of the molecular plane and z pointing along the molecular 2-fold rotation axis. [d] Includes relativistic effects.

of the cartesian d's were not included but p-symmetry components of the cartesian f's were kept. This basis contained 120 contracted functions and is referred to as the "extended ETG" basis in the tables.

The tables also contain two other nonstandard basis sets for comparison. One is a minimal, atomic orbital basis formed from the Hartree–Fock orbitals of the atoms computed in a large Gaussian basis with spin and symmetry constraints. The second is a split valence plus polarization basis which was formed from the natural orbitals of the atoms by including the first shell of correlating orbitals. For hydrogen, the supplemental orbitals were taken from an sp-limit calculation on $H_2{}^+$ and reproduce exactly the hydrogen atom and $H_2{}^+$ orbitals of the uncontracted basis.

As Table I shows, the Dunning basis sets give better absolute energies than the Pople basis sets of comparable size. The nonstandard bases, however are much better than any of the standard sets. The atomic orbital set is 1300 mH below the STO-3G energy. The atomic NO set is 10 mH below other split-valence sets. Even so, the extended set with CI is still 220 mH above the experimental energy (about 80 mH of this is core correlation and 30 mH is relativistic energy).

It is customary to assume, based on extensive experience, that energy differences will be better than absolute energies. The energy of atomization computed with STO-3G approaches most closely to the SCF limit, but a large piece of this energy is clearly due to basis set superposition error (BSSE). In spite of the small BSSE of the AO basis, it gives the worst energy of atomization. Only bases at the split-valence plus polarization level are capable of describing the change in size and shape of the atomic orbitals in the molecule, so only this level of calculation can obtain reasonable atomization energies for the right reasons. When correlation energy is recovered via CI, the difference between large and small basis sets generally increases.

All of the Pople bases have larger BSSE than the Dunning bases. Generally the Dunning bases give more accurate energies of atomization at both the SCF and CI levels than the Pople bases with comparable numbers of contracted functions. The vertical ionization and excitation energies with both classes of bases are comparable in accuracy. Notice that no $^1A_1$ excited state was computed because the first excited state of this symmetry (the V state) is a $2b_2 \rightarrow 3b_2$ (3py) Rydberg excitation which cannot be described by any of these basis sets. As is well known, any attempt to compute this state with a valence-only basis gives a $1b_1 \rightarrow 2b_1$ state above the $2b_2$ ionization limit.

Tables II and III give some other properties computed from these basis sets. Because these properties tend to be first-order in the error in the wavefunction while the energy is second-order, they are a better indication of wave function quality. Again the STO-3G and AO bases are notable for their poor results. All other bases give a reasonable dipole moment although the Dunning bases are somewhat better than the Pople bases. The quadrupole moment results are much more erratic, and none of the standard bases produce an accurate result although, again, the Dunning bases are

TABLE II. Selected Origin-Centered Properties[a] for Formaldehyde ($^1A_1$) at the Experimental Geometry

basis	$\langle \mu z \rangle$	θ_{xx}	θ_{yy}	$\langle r^2 \rangle_e$	$\langle p^4 \rangle_e$
Dunning/Hay	−1.224	−0.015	0.433	61.119	48825.8
SV	−1.206	0.256	0.268	61.551	48825.6
Dunning/Hay	−1.124	−0.181	0.219	60.453	48775.2
SVP	−0.963	−0.013	0.138	60.708	48802.8
Dunning DZP	−1.124	−0.182	0.218	60.452	48775.2
	−0.960	−0.016	0.137	60.701	48825.5
STO-3G	−0.600	0.287	0.220	58.641	32477.4
	−0.458	0.396	0.165	58.857	32488.9
3-21G	−1.038	0.042	0.258	60.856	39452.8
	−0.798	0.302	0.108	61.197	39469.6
6-31G	−1.184	0.057	0.287	61.067	48164.3
	−0.968	0.331	0.141	61.449	48164.0
6-31G**	−1.085	−0.061	0.093	60.479	48030.2
	−0.919	0.109	0.027	60.674	48058.5
MIDI-4	−1.167	0.096	0.253	61.015	43868.3
	−0.957	0.351	0.109	61.412	43871.4
HF AO	−0.785	−0.107	0.405	65.508	50373.3
minimal	−0.642	0.064	0.318	65.791	50385.3
atomic NO	−1.106	−0.291	0.242	60.457	50363.2
SVP	−0.939	−0.139	0.176	61.598	50375.6
(s,p,d,f)	−1.124	−0.155	0.239	60.749	50422.5
extended	−0.987	−0.041	0.177	60.813	50474.6
exptl	−0.917 ± 0.008 (C^+O^-)	1.35 ± 1.0	−0.45 ± 0.5		

[a] Properties are in atomic units and are measured with respect to the center-of-mass. For each basis set entry SCF values are given in the first row, followed by CI values in the second row. The properties (in atomic units) are defined by $\langle \mu z \rangle = \langle \sum q_i z_i \rangle$, $\theta_{xx} = 0.5\langle \sum q_i(3x_i^2 - r_i^2) \rangle$ where the sum runs over nuclei and electrons; $\langle r^2 \rangle_e = \langle \sum r_i^2 \rangle$, $\langle p^4 \rangle_e = \langle \sum p_i^4 \rangle$ where the sum runs over electrons only.

better than the Pople ones. The dipole and quadrupole moments weight regions far from the molecule. Consequently diffuse supplemental basis functions are required for reliability. Also, since the quadrupole moment measures the $l = 2$ component of the density, diffuse polarization functions of both d- and f-type are expected to be important.

The field gradient q and its asymmetry are usually regarded as difficult to compute. Except for the minimum basis sets, q was easily obtained for oxygen within 10%. The hydrogen results were somewhat better. The carbon results, on the other hand, show no clear convergence. An operator such as q, which has $l = 2$ and weights the core region heavily, requires higher l functions in the core and valence regions. Thus, following Sternheimer,[117] one would expect that d functions would be needed in the 1s core region and up to f functions would be needed in the valence region before the calculation could have predictive ability for q.

Difficulties

While the above comparison is instructive, not all molecules are as well behaved as the closed shell ground state of formaldehyde. The literature contains numerous references to classes of compounds and properties which present much more severe tests of basis set quality. A small sampling will be presented next.

The overwhelming majority of quantum chemical calculations are done in order to find molecular geometries and relative energies. Therefore, a feeling for the kind of reliability to be expected for these two properties is of great importance. As discussed in the section describing individual basis sets, the structure of most molecules is predicted quite well with any of the standard split valence or split valence polarized basis sets. In a review article on the status of ab initio techniques for determining molecular geometries Schafer[118] has reported that the results of hundreds of calculations indicate that absolute accuracy to ±0.01 Å in bond lengths and ±1° in bond angles can be achieved through the use of empirical correction factors. Although no comprehensive table exists which shows the accuracy in bond lengths and angles obtained with all commonly used basis sets for the same set of mole-

TABLE III. Selected Atom-Centered Properties[1] for Formaldehyde ($^1A_1$) at the Experimental Geometry

	oxygen				carbon				hydrogen			
basis	$\langle \delta \rangle_e$	$\langle 1/r \rangle_e$	q	η	$\langle \delta \rangle_e$	$\langle 1/r \rangle_e$	q	η	$\langle \delta \rangle_e$	$\langle 1/r \rangle_e$	q	η
Dunning/Hay	295.6	25.470	−2.608	0.475	119.8	19.067	0.490	0.383	0.415	6.259	0.287	0.007
SV	295.9	25.442	−2.319	0.486	119.8	19.093	0.356	0.628	0.424	6.275	0.289	0.006
Dunning/Hay	295.4	25.478	−2.343	0.515	119.6	19.065	0.576	0.080	0.413	6.279	0.256	0.013
SVP	295.5	25.451	−2.123	0.528	119.6	19.099	0.459	0.207	0.419	6.291	0.258	0.016
Dunning DZP	295.4	25.475	−2.342	0.515	119.7	19.066	0.577	0.077	0.413	6.279	0.256	0.013
	295.6	25.451	−2.122	0.529	119.7	19.101	0.460	0.202	0.419	6.291	0.259	0.002
STO-3G	193.6	25.143	−2.443	0.252	77.4	18.912	0.161	0.364	0.365	6.333	0.277	0.000
	193.7	25.116	−2.358	0.271	77.4	18.933	0.203	0.463	0.375	6.352	0.278	0.009
3-21G	233.8	25.321	−2.223	0.641	94.0	18.972	0.429	0.225	0.367	6.262	0.279	0.009
	233.8	25.325	−2.030	0.667	94.0	19.011	0.315	0.441	0.374	6.280	0.276	0.022
6-31G	291.6	25.451	−2.428	0.546	118.4	19.067	0.510	0.226	0.423	6.263	0.280	0.000
	291.5	25.435	−2.187	0.552	118.3	19.097	0.388	0.415	0.432	6.279	0.279	0.012
6-31G**	290.9	25.446	−2.219	0.562	118.2	19.068	0.589	0.014	0.424	6.283	0.247	0.012
	290.9	25.429	−2.026	0.556	118.3	19.109	0.478	0.132	0.430	6.295	0.248	0.002
MIDI-4	261.6	25.401	−2.481	0.553	105.9	19.023	0.504	0.236	0.375	6.253	0.286	0.004
	261.5	25.387	−2.240	0.559	105.9	19.056	0.381	0.423	0.380	6.266	0.285	0.008
HF AO	307.8	25.401	−3.441	0.557	124.7	18.857	0.207	0.831	0.249	6.035	0.381	0.004
minimal	307.9	25.376	−3.243	0.573	124.7	18.875	0.261	0.593	0.261	6.052	0.382	0.009
atomic NO	307.6	25.488	−2.521	0.439	125.0	19.072	0.489	0.159	0.401	6.272	0.269	0.010
SVP	307.6	25.475	−2.363	0.462	125.0	19.105	0.385	0.327	0.403	6.279	0.271	0.002
(s,p,d,f)	308.9	25.478	−2.272	0.642	122.7	19.077	0.672	0.075	0.452	6.280	0.245	0.007
extended	309.1	25.486	−2.133	0.627	122.7	19.121	0.578	0.020	0.456	6.288	0.243	0.020
exptl			−2.194	0.695							0.261	0.018

[a] Properties are in atomic units. For each basis set entry SCF values are given in the first row, followed by CI values in the second row. The properties (in atomic units) are defined as $\langle \delta_x \rangle_e = \langle \sum \delta(r_{ix}) \rangle$, $\langle 1/r_x \rangle_e = \langle \sum 1/r_{ix} \rangle$ where the sum runs over all electrons for a given nucleus x; q_x = the field gradient at nucleus x, *ex.* for oxygen $q_0 = q_{yy} = \langle \sum qi(3x_i^2 - r_i^2)/r_i^5 \rangle$ where the sum runs over all electrons and nuclei except oxygen, η_x = the asymmetry parameter.

cules, some conclusions can be drawn from the fragmentary information which is available.

In an appendix for the Japanese Quantum Chemistry Literature Data Base Iwata[119] states that the average deviation in bond lengths for minimal basis sets is +0.05 Å and ±10° in bond angles at the SCF level. If we stay at the SCF level but increase the basis set to the split valence or double zeta level the average deviations drops to −0.02 Å and ±5°. Further improvements in the basis set (polarization functions, triple zeta valence space) do not produce concomitant improvements in bond lengths although deviations in bond angles do improve slightly to ±3°. The expected improvements finally show up only when electron correlation effects are considered. Deviations then drop to the ±0.003–0.005 Å and ±1° level. Care must also be taken in defining the "experimental" geometries when differences on the order of 0.002 Å are in question.

Even here it must be remembered that part of the excellent agreement with experiment arises from a cancellation of errors due to basis set incompleteness and lack of full correlation recovery. For instance, if one goes from a 6-31G* basis to a 6-311G** basis, both at the MP3 level, the agreement gets worse (bonds get shorter). Conversely, if one goes from MP3 to MP4 with the same 6-31G* basis agreement also worsens (bonds get longer). Many occasional users of ab initio programs mistakenly equate longer run times with greater accuracy.

Instances of poor agreement between SCF and experimental geometries are much more prevalent in open shell systems. Based on a relatively small sampling of di- and triatomics (many of which were hydrides) Farnell, Pople, and Radom[120] reported an accuracy for RHF, UHF, and UMP2 open-shell techniques comparable to the accuracy in closed-shell molecules with the same 6-31G* basis. However, they warned that in cases where the degree of spin contamination in the UHF wave function becomes large the RHF results are preferred and the UHF wave function would probably provide a poor starting point for the UMP perturbation theory expansion.

In spite of the optimistic flavor of the above article there are numerous studies of radicals and diradicals in the literature which report difficulties. For these species the potential energy surfaces are often quite flat. Subtle basis set and/or correlation effects can qualitatively change the geometries obtained by ab initio methods. Some of the difficulties manifest themselves as artifactual symmetry breaking phenomena,[121] i.e., certain methods result in unsymmetrical wave functions at symmetrical geometries.

Other problems can arise from physically unreasonable solutions of the Hartree–Fock equations. For example, in SiO^+ ($^2\Sigma^+$) an extended (s,p,d) basis set SCF wave function[122] placed the unpaired electron primarily on oxygen in spite of strong experimental evidence to the contrary. The results from a large MCSCF wave function were better. While the dominant configuration possessed the same orbital occupancy as the HF configuration the singly occupied orbital was delocalized over both atoms.

Even closed-shell ground states of molecules composed of first-row atoms can be difficult. F_2 and F_2O_2[123,124] are excellent examples where polarized basis set SCF wave functions predicted bond lengths which differed from experiment by more than 0.1 Å. Extensively correlated wave functions are necessary to obtain good results for these molecules. Another example is the transition metal dimer Cr_2 whose experimental bond length is 1.68 Å, to be compared with an (s,p,d,f,g) SCF[125] result of 1.45 Å. Enough other examples are known that care is always warranted in computing geometries with small to medium sized basis sets at the Hartree–Fock level.

TABLE IV. Calculated Total Energies (hartrees) and Electron Affinities (eV) of the Oxygen Atom using Spin- and Symmetry-Restricted SCF Wave Functions

basis sets	neutral $E(^3P)$	anion $E(^2P)$	EA
Basis Sets *without* Extra Diffuse Functions			
STO-3G	−73.8041	−73.4269	−10.26
3-21G	−74.3918	−74.2715	−3.27
6-31G	−74.7769	−74.7144	−1.70
Dunning DZ	−74.7988	−74.7538	−1.22
Basis Sets *with* Extra Diffuse Functions			
3-21+G	−74.4093	−74.3964	−0.35
Dunning DZ+	−74.7993	−74.7779	−0.58
extended	−74.8091	−74.7894	−0.54
numerical HF	−74.8091	−74.7894	−0.54
exptl			1.46

For hydrogen-bonded systems the weakness of the bond causes problems in that the basis set superposition error can be a sizeable percentage of the interaction energy. As discussed previously for van der Waals complexes the BSSE produces a distortion in the optimal geometry. The 3-21G SCF O–O distance in the water dimer is 2.80 Å compared to the experimental value of 2.98 Å. Even with a moderately large [5s,4p,1d/3s,1p] basis the SCF BSSE amounts to 20% of the binding energy at the equilibrium bond length and 60% at the SD–CI level.[126] An improvement to the counterpoise correction scheme, intended specifically to improve the description of hydrogen bonded systems, has been suggested by Loushin et al.[127]

Electron affinities (EA's) have a well-deserved reputation for being difficult to compute via ab initio techniques. This difficulty can be viewed as arising from two distinct problems. The first problem, which appears even at the SCF level has to do with an increase in the diffuseness of the orbital receiving the extra electron. As can be seen in Table IV, standard basis sets, which lack sufficiently diffuse functions, do very poorly when compared to the numerical HF or converged basis set limit for the oxygen atom.

Table IV also illustrates the second problem encountered when trying to accurately compute the energy difference between a molecule (or atom) and its negative ion. While the extra electron is experimentally found to be bound by 1.46 eV, at the Hartree–Fock level the anion is 0.54 eV *above* the neutral. Thus, the correlation correction for the oxygen atom EA amounts to +2 eV. The magnitude of the correction is not unique to electron affinities. Similar sized corrections have been shown for ionization potentials. The convergence of the correlation correction as a function of the number and types of polarization functions, shown in Table V, is discouragingly slow if one believes that in the complete basis set limit the ab initio result will agree with experiment. Full CI calculations on (O,O^-)[128] and (F,F^-)[129] using moderate sized basis sets are being used

TABLE V. Electron Affinities (eV) for the Oxygen Atom as a Function of the Number and Type of Higher L Basis Functions at the CI Level[a]

#d's	#f's	#g's	*EA*
0	0	0	1.01
1	0	0	0.96
2	0	0	1.11
3	0	0	1.16
4	0	0	1.19
4	1	0	1.25
4	2	0	1.30
4	2	1	1.32

[a] The CI calculations were of the multireference singles and doubles variety. The reference space consisted of single and double excitations in the (s,p) space.

to unambiguously indicate the high percentage of the correlation energy which must be recovered to accurately treat electron affinities.

We have already discussed the basis set requirements for certain properties. Another set of properties which has received attention lately and for which the basis set requirements have proven discouragingly large is the set of hyperpolarizabilities. 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

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

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. In work on H_2O[130] and HF[131] the need for quite diffuse functions has been demonstrated. In the former case, 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,6d/5s,3p] basis set at the MBPT level. For HF an even larger [5s,3p,4d,2f/5s,3p] basis, capable of yielding 92% of the valence correlation energy at the MBPT (4) level, was used. In spite of the number of basis functions the predicted values of the β hyperpolarizability was only 50–70% of the experimental value in water. Only slightly better results were found for HF. This presents a conceptual, as well as computational, problem since in the limit of a complete set of diffuse basis functions the molecule will be predicted to autoionize and bound state methods are not appropriate.

Conclusion

The sustained high publication rate for papers introducing new Gaussian basis sets or analyzing the performance of existing sets attests to their central role in current quantum chemical methodologies. Indeed, whether one finds the topic of basis sets intrinsically interesting or not, no researcher involved in performing quantum chemical calculations or in understanding the significance of someone else's calculations can afford to be totally ignorant of the common basis sets and their limitations. Furthermore, as the impact of ab initio predictions continue to grow and influence more areas of chemistry the need to know will also grow.

The choice of basis set is inextricably tied to the hardware and software capabilities which can be devoted to a problem. At present quantum chemistry programs are running across a vast spectrum of computers ranging from the latest American and Japanese supercomputers all the way down to personal computers. Machines in a class with the 256 Mword, several hundred million floating point operations per second Cray-2, are just beginning to be explored. It is, however, evident that such computers will enable much larger basis sets to be used than have been previously possible. At the same time, advances in rotating disk technology and declines in cost have permitted individual chemistry departments to possess gigabytes worth of storage capacity. So that even at the local level calculations are being done that would have been impossible only 10 years ago.

In the latter half of the 1980s calculations employing 100 or so basis functions have become commonplace. "Large" calculations, in terms of numbers of basis functions, fall into the 200–300 function range. Since this range is independent of the number of atoms you can either do very extended basis set calculations on small systems or smaller basis set calculations on larger systems. For the most part the use of minimal basis sets has been relegated to those molecular systems with very many atoms (>50) or to extensive scans of potential energy surfaces. The continuing development of relativistic effective core potentials is opening up rows of the periodic table that are nearly impossible to treat with nonrelativistic, all-electron methods.

The mounting evidence seems to suggest that small split valence (e.g., 3-21G or Dunning/Hay SV) or split valence plus polarization (e.g. 6-31G*) basis sets perform adequately in SCF or correlated geometry optimizations of closed shell organic compounds, but that very extended sets with flexible valence spaces and several sets of polarization functions are needed for energy differences accurate to 2–5 kcal/mol or 1-electron properties accurate to within 5%. Such calculations represent a challenge even with state-of-the-art facilities. The desire to choose the most efficient basis set to obtain a desired answer will undoubtedly motivate continuing research in this area.

References

(1) Pople, J. A.; Seeger, R.; Krishnan, R. *Int. J. Quant. Chem. Symp.* **1977**, *11*, 149.
(2) Huzinaga, S. *Comp. Phys. Reports* **1985**, *2*, 279.
(3) Wilson, S. In *Methods in Computational Molecular Physics* Diercksen, G. H., Wilson, S., Eds. Dordrecht, The Netherlands, Reidel: 1983.
(4) Boys, S. F. *Proc. R. Soc. London, A* **1950**, *A200*, 542.
(5) Regier, P. E.; Fisher, J.; Sharma, B. S.; Thakkar, A. J. *Int. J. Quantum Chem.* **1985**, *28*, 429.
(6) Carsky, P.; Urban, M. *Ab Initio Calculations*; Springer Verlag: Heidelberg, 1980.
(7) Rosenberg, B. J.; Shavitt, I. *J. Chem. Phys.* **1975**, *63*, 2162.
(8) Clementi, E. *IBM J. Res. and Dev.* **1965**, *9*, 2.
(9) Whitten, J. L. *J. Chem. Phys.* **1966**, *44*, 359.
(10) Clementi, E.; Davis, D. R. *J. Comput. Phys.* **1967**, *1*, 223.
(11) Clementi, E.; Veillard, A. QCPE Program 92, **1966.**
(12) Hunt, W. J.; Goddard, W. A., III. *Chem. Phys. Lett.* **1969**, *3*, 414.
(13) Huzinaga, S. *J. Chem. Phys.* **1965**, *42*, 1293.
(14) Davidson, E. R.; McMurchie, L. E.; Day, S. *J. Chem. Phys.* **1981**, *74*, 739.
(15) Davidson, E. R.; Feller, D. *Chem. Phys. Lett.* **1984**, *104*, 54.
(16) Ruedenberg, K.; Raffenetti, R. C.; Bardo, R. D. *Energy, Structure and Reactivity, Proceedings of the 1972 Boulder Conference*; Wiley: New York, 1973.
(17) Bardo, R. D.; Ruedenberg, K. *J. Chem. Phys.* **1974**, *60*, 918.
(18) Reeves, C. M. *J. Chem. Phys.* **1963**, *39*, 1.
(19) Feller, D.; Ruedenberg, K. *Theor. Chim. Acta* **1979**, *71*, 231.
(20) Schmidt, M. W.; Ruedenberg, K. *J. Chem. Phys.* **1979**, *71*, 3951.

(21) Feller, D.; Davidson, E., to be published.
(22) Bardo, R. D.; Ruedenberg, K. *J. Chem. Phys.* **1974**, *60*, 918.
(23) Huzinaga, S.; Klobukowski, M.; Tatewaki, H. *Can. J. Chem.* **1985**, *63*, 1812.
(24) Huzinaga, S.; Klobukowski, M. *Chem. Phys. Lett.*, in press.
(25) Klobukowski, M.; Huzinaga, S. *J. Chem.Phys.* **1986**, *84*, 323.
(26) Silver, D. M.; Nieuwpoort, W. C. *Chem. Phys. Lett.* **1978**, *57*, 421.
(27) Silver, D. M.; Wilson, S.; Nieuwpoort, W. C. *Int. J. Quantum Chem.* **1978**, *14*, 635.
(28) Klahn, Bruno. *J. Chem. Phys.* **1985**, *83*, 5754.
(29) Clementi, E.; Corongiu, G. *Chem. Phys. Lett.* **1982**, *90*, 359.
(30) Feller, D.; Davidson, E. R. *J. Chem. Phys.* **1985**, *82*, 4135.
(31) Jankowski, K.; Becherer, R.; Scharf, P.; Schiffer, H.; Ahlrichs, R. *J. Chem. Phys.* **1985**, *82*, 1413.
(32) Becherer, R.; Ahlrichs, R. *Chem. Phys.* **1985**, *99*, 389.
(33) Poirier, R.; Kari, R. *Can. J. Chem.* **1978**, *56*, 543.
(34) Roos, B.; Siegbahn, P. *Theor. Chim. Acta* **1970**, *17*, 208.
(35) Magnusson, E.; Schaefer, H. F., III. *J. Chem. Phys.* **1985**, *83*, 5721.
(36) Roos, B. O.; Sadlej, A. J. *Chem. Phys.* **1985**, *94*, 43.
(37) Rothenberg, S.; Schaefer, H. F., III. *J. Chem. Phys.* **1971**, *54*, 2765.
(38) Butscher, W.; Shih, S.; Buenker, R. J.; Peyerimhoff, S. D. *Chem. Phys. Lett.* **1977**, *52*, 457.
(39) Neisius, D.; Verhaegen, G. *Chem. Phys. Lett.* **1979**, *66*, 358.
(40) Lawlor, L.; Vasudevan, K.; Grein, F. *Chem. Phys. Lett.* **1980**, *75*, 79.
(41) Mach, P.; Kysel, O. *J. Comput. Chem.* **1985**, *6*, 312.
(42) Bauschlicher, C. W. ICASE Report No. 80-9 CI Superposition Error Caused by Bond Functions; 1980, NASA Langley Research Center, Hampton, VA.
(43) Wright, J. S.; Williams, R. J. *J. Chem. Phys.* **1983**, *78*, 5264.
(44) Wright, J. S.; Buenker, R. J. *Chem. Phys. Lett.* **1984**, *106*, 570.
(45) Bauschlicher, C. W. *Chem. Phys. Lett.* **1985**, *122*, 572.
(46) Raffenetti, R. C. *J. Chem. Phys.* **1973**, *58*, 4452.
(47) Bardo, R. D.; Ruedenberg, K. *J. Chem. Phys.* **1974**, *60*, 918.
(48) Cheung, L. M. Ph.D. Dissertation, Iowa State University, 1975.
(49) Examples of programs which do handle the general contraction scheme efficiently are: BIGMOLLI by R. Raffenetti, QCPE 11 (1977), and ARGOS by R. M. Pitzer.
(50) Mulliken, R. S. *J. Chem. Phys.* **1962**, *36*, 3428.
(51) Wilson, S. *J. Mol. Struct.* (*Theochem*) **1986**, *135*, 135.
(52) Maroulis, G.; Sana, M.; Leroy, G. *J. Mol. Struct.* (*Theochem*) **1985**, *122*, 269.
(53) Maroulis, G.; Sana, M.; Leroy, G. *J. Mol. Struct.* (*Theochem*) **1985**, *121*, 69.
(54) Kari, R. E.; Mezey, P. G.; Csizmadia, I. G. *J. Chem. Phys.* **1977**, *63*, 581.
(55) Mezey, P. G.; Kari, R. E.; Csizmadia, I. G. *J. Chem. Phys.* **1977**, *66*, 964.
(56) Andzelm, J.; Klobukowski, M.; Radzio-Andzelm, E. *J. Comput. Chem.* **1984**, *5*, 146.
(57) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(58) Johnson, A.; Kollman, P.; Rothenberg, S. *Theor. Chim. Acta* **1973**, *29*, 49.
(59) Morokuma, K.; Kitaura, K. In *Chemical Applications of Atomic and Molecular Electrostatic Potentials* Politzer, P., Ed.; Plenum: New York, 1981.
(60) Diercksen, G. H. F.; Kello, V.; Sadlej, A. J. *Chem. Phys.* **1986**, *103*, 55.
(61) Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *82*, 2418.
(62) Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *84*, 4113.
(63) Gutowski, M.; van Lenthe, J. H.; Verbeek, J.; van Duijneveldt, F. B.; Chalasinki, G. *Chem. Phys. Lett.* **1985**, *124*, 370.
(64) Christiansen, P. A.; Lee, Y. S.; Pitzer, K. S. *J. Chem. Phys.* **1979**, *71*, 4445.
(65) Stevens, W. J.; Basch, H.; Krauss, M. *J. Chem. Phys.* **1984**, *81*, 6026.
(66) Topiol, S.; Osman, R. *J. Chem. Phys.* **1980**, *73*, 5191.
(67) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270.
(68) Wadt, W. R.; Hay, P. J. *J. Chem. Phys.* **1985**, *82*, 284.
(69) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.
(70) Jansky, J.; Bartram, R. H.; Rossi, A. R.; Corradi, G. *Chem. Phys. Lett.* **1986**, *124*, 26.
(71) Rothenberg, S.; Schaefer, H. F., III. *J. Chem. Phys.* **1970**, *53*, 3014.
(72) Bachelet, G. B.; Hamann, D. R.; Schuler, M. *Phys. Rev. B* **1982**, *26*, 4199.
(73) Rappe, A. K.; Smedley, T. A.; Goddard, W. A., III. *J. Chem. Phys.* **1981**, *85*, 1662.
(74) Poirier, R.; Kari, R.; Csizmadia, I. G. *Handbook of Gaussian Basis Sets*; Elsevier Science: New York, 1985.
(75) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657.
(76) Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1970**, *52*, 2769.
(77) Pietro, W. J.; Levi, B. A.; Hehre, W. J.; Stewart, R. F. *Inorg. Chem.* **1980**, *19*, 2225.
(78) Pietro, W. J.; Blurock, E. S.; Hout, R. S.; Hehre, W. J.; DeFrees, D. J.; Stewart, R. F. *Inorg. Chem.* **1981**, *20*, 3650.
(79) Pietro, W. J.; Hehre, W. J. *J. Comput. Chem.* **1983**, *4*, 241.
(80) Pople, J. A. In *Modern Theoretical Chemistry*, Schaefer, H. F., III, Ed.; Plenum: New York, 1976, Vol. 4.
(81) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. *J. Am. Chem. Soc.* **1979**, *101*, 2550.
(82) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939.
(83) Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J. *J. Am. Chem. Soc.* **1982**, *104*, 2797.
(84) Pietro, W. J.; Francl, M. M.; Hehre, W. J.; Pople, J. A.; Binkley, J. S. *J. Am. Chem. Soc.* **1982**, *104*, 5048.
(85) Clark, T.; Chandrasekhar, J.; Spitznagel, G.; van Rague Schleyer, P. *J. Comput. Chem.* **1983**, *4*, 294.
(86) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.
(87) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.
(88) Hehre, W. J.; Latham, W. A. *J. Chem. Phys.* **1972**, *56*, 5255.
(89) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654.
(90) Nobes, R. H.; Rodwell, W. R.; Radom, L. *J. Comput. Chem.* **1982**, *3*, 561.
(91) DeFrees, D. J.; Levi, B. A.; Binkley, J. S.; Pople, J. A. *J. Am. Chem. Soc.* **1979**, *101*, 4085.
(92) Chandrasenkhar, J.; Andrade, J. G.; van Rague Schleyer, P. *J. Am. Chem. Soc.* **1981**, *103*, 5609.
(93) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.
(94) DeFrees, D. J.; Binkley, J. S.; McLean, A. D. *J. Chem. Phys.* **1984**, *80*, 3720.
(95) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265.
(96) Dunning, T. H., Jr. *J. Chem. Phys.* **1970**, *53*, 2823.
(97) Dunning, T. H., Jr.; Hay, P. J. In *Modern Theoretical Chemistry*; Plenum Press, New York, 1977; Vol. 2.
(98) McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639.
(99) Basch, H.; Hornback, C. J.; Moskowitz, J. W. *J. Chem. Phys.* **1969**, *51*, 1311.
(100) Wachters, A. J. H. *J. Chem. Phys.* **1970**, *52*, 1033.
(101) Roos, B.; Veillard, A.; Vinot, G. *Theor. Chim. Acta* **1971**, *20*, 1.
(102) Hay, P. J. *J. Chem. Phys.* **1977**, *66*, 4377.
(103) McClean, A. D.; Liu, B. *Chem. Phys. Lett.* **1983**, *101*, 144.
(104) Walch, S. P.; Bauschlicher, C. W. *Chem. Phys. Lett.* **1984**, *105*, 171.
(105) Huzinaga, S. *J. Chem. Phys.* **1977**, *66*, 4245.
(106) Walch, S. P.; Bauschlicher, C. W., Jr.; Nelin, C. J. *J. Chem. Phys.* **1983**, *79*, 3600.
(107) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1970**, *52*, 5001.
(108) Tavouktsoglou H.; Huzinaga, S. *J. Comput. Chem.* **1980**, *3*, 205.
(109) Tavouktsoglou H.; Huzinaga, S. *J. Chem. Phys.* **1979**, *71*, 4339.
(110) Veillard, A. *Theor. Chim. Acta* **1968**, *12*, 405.
(111) Roos, B.; Siegbahn, P. *Theor. Chim. Acta* **1970**, *17*, 209.
(112) Tatewaki, H.; Huzinaga, S. *J. Chem. Phys.* **1979**, *71*, 4339.
(113) Tatewaki, H.; Huzinaga, S. *J. Chem. Phys.* **1980**, *72*, 399.
(114) Stromberg, A.; Gropen, O.; Wahlgren, U. *J. Comput. Chem.* **1983**, *4*, 181.
(115) Oka, T. *J. Phys. Soc. Jpn.* **1960**, *15*, 2274.
(116) Takagi, K.; Oka, T. *J. Phys. Soc. Jpn.* **1963**, *18*, 1174.
(117) Sternheimer, R. N. *Phys. Rev.* **1963**, *130*, 1423.
(118) Schafer, L. *J. Mol. Struct.* (*Theochem*) **1983**, *100*, 51.
(119) Iwata, S. In *Quantum Chemistry Literature Data Base*; Ohno, K., Morokuma, K., Eds.; Elsevier: Amsterdam, 1982.
(120) Farnell, L.; Pople, J. A.; Radom, L. *J. Phys. Chem.* **1983**, *87*, 79.
(121) Davidson, E. R.; Borden, W. T. *J. Phys. Chem.* **1983**, *87*, 4783.
(122) Knight, L. B., Jr.; Ligon, A.; Woodward, R. W.; Feller, D.; Davidson, E. R. *J. Am. Chem. Soc.* **1985**, *107*, 2857.
(123) Lucchese, R. R.; Schaefer, H. F., III.; Rodwell, W. R.; Radom, L. *J. Chem. Phys.* **1978**, *68*, 2507.
(124) Alrichs, R.; Taylor, P. R. *Chem. Phys.* **1982**, *72*, 287.

(125) McClean, A. D.; Liu, B. *Chem. Phys. Lett.* **1983**, *101*, 144.
(126) Baum, J. O.; Finney, J. L. *Mol. Phys.* **1985**, *55*, 197.
(127) Loushin, S. K.; Liu, S.; Dykstra, C. E., to be published.
(128) Bauschlicher, C. W.; Langhoff, S. R.; Partridge, H.; Taylor, P. R., submitted for publication in *J. Chem. Phys.*
(129) Bauschlicher, C. W.; Taylor, P. R., to be published.
(130) Purvis, G. D.; Bartlett, R. J. *Phys. Rev. A* **1981**, *23*, 1594.
(131) Sekino, H.; Bartlett, R. J. *J. Chem. Phys.* **1986**, *84*, 2726.