

badly) such calculations compare with experiment. We will restrict our calculations to H<sub>2</sub>, N<sub>2</sub>, and CO and molecules in the ten-electron series CH<sub>4</sub>, NH<sub>3</sub>, H<sub>2</sub>O, and FH. For each of these we will use a hierarchy of well-defined basis sets. Prior to describing the results of such calculations, however, we need to discuss the general question of polyatomic basis sets and describe the specific basis sets we will be using.

### 3.6 POLYATOMIC BASIS SETS

There are probably as many basis sets defined for polyatomic calculations as there are quantum chemists. The choice of a basis set is not nearly the black art, however, that it may first appear. In our sample calculations we will use a reasonably well-defined hierarchy of basis sets starting with a minimal STO-3G basis and proceeding through the 4-31G basis, which effectively doubles the number of functions, the 6-31G\* basis, which adds *d*-type functions to heavy atoms C, N, O, and F and finally, the 6-31G\*\* basis, which, in addition to *d*-type functions for heavy atoms, adds *p*-type functions to hydrogen. By performing electronic structure calculations on a small variety of molecules, using this hierarchy of basis sets, it is possible to gain some insight into the size and characteristics of a basis set needed to obtain a given level of calculational accuracy.

The above basis sets have been introduced by Pople and collaborators (see Hehre *et al.* in Further Reading at the end of this chapter) and have been used extensively, by a number of workers for calculations on a large variety of molecules. Apart from a few instances, all the calculations in this book use the STO-3G, 4-31G, 6-31G\*, and 6-31G\*\* hierarchy of basis sets. By restricting our example calculations to a very limited set of molecules and the above basis sets, we are attempting to illustrate in a systematic way how specific attributes of a basis set affect calculated quantities. We are not attempting to provide a general review of current calculations. Such a review would be out of date very quickly. While the basis sets we use are not necessarily optimum, and may themselves be out of date shortly, they do have characteristics that can be used to illustrate all basis sets.

Our purpose in this section is to explicitly define the STO-3G, 4-31G, 6-31G\*, and 6-31G\*\* basis sets that we will be using in this and subsequent chapters. In the process, however, we will describe attributes that are characteristic of most of the basis sets that are in current use, and we will introduce some of the notation and some of the mechanics of defining and choosing a basis set. In particular, we first present a general treatment of contraction.

#### 3.6.1 Contracted Gaussian Functions

In Subsection 3.5.1, when we defined the 1s STO-3G basis set of our model calculations, we indicated some of the ideas pertinent to the concept of contraction. We review these briefly. There are two main considerations in the

choice of a basis. The first is that one desires to use the most efficient and accurate functions possible, in the sense that the expansion

$$\psi_i = \sum_{\mu=1}^K C_{\mu i} \phi_{\mu} \quad (3.282)$$

will require the fewest possible terms for an accurate representation of the molecular orbitals  $\psi_i$ . From this consideration, Slater functions are better than Gaussian functions. The second consideration in the choice of a basis set is the speed of two-electron integral evaluation. Here Gaussian functions have the advantage. By using a basis set of *contracted Gaussian functions* one can in a sense have one's cake and eat it too. In this procedure, one lets each basis function be a fixed linear combination (contraction) of Gaussian functions (primitives). Prior to a calculation one chooses the exponents of the primitives and the contraction coefficients so as to lead to basis functions with desired qualities. The contracted basis functions might be chosen to approximate Slater functions, Hartree-Fock atomic orbitals, or any other set of functions desired. Integrals involving such basis functions reduce to sums of integrals involving the primitive Gaussian functions. Even though many primitive integrals may need to be calculated for each basis function integral, the basis function integrals will be rapidly calculated provided the method of computing primitive integrals is very fast.

To avoid confusion with a multitude of  $\phi$ 's we will use the symbol  $g$  here for a normalized Gaussian function. A contraction thus has the form.

$$\phi_{\mu}^{\text{CGF}}(\mathbf{r} - \mathbf{R}_A) = \sum_{p=1}^L d_{p\mu} g_p(\alpha_{p\mu}, \mathbf{r} - \mathbf{R}_p) \quad (3.283)$$

where  $\alpha_{p\mu}$  and  $d_{p\mu}$  are the contraction exponents and coefficients and  $L$  is the length of the contraction. The normalized Gaussian primitive functions are of the  $1s, 2p, 3d, \dots$  type,

$$g_{1s}(\alpha, \mathbf{r}) = (8\alpha^3/\pi^3)^{1/4} e^{-\alpha r^2} \quad (3.284)$$

$$g_{2p_x}(\alpha, \mathbf{r}) = (128\alpha^5/\pi^3)^{1/4} xe^{-\alpha r^2} \quad (3.285)$$

$$g_{3d_{xy}}(\alpha, \mathbf{r}) = (2048\alpha^7/\pi^3)^{1/4} xy e^{-\alpha r^2} \quad (3.286)$$

The simplifications that occur for integral evaluation using these functions do not appear for the  $2s, 3p, \dots$  Gaussians, and so any basis function of  $s$  symmetry, for example a  $2s$  or  $3s$  Slater function, will be expanded in only  $1s$  Gaussians, with similar restrictions on the other symmetry types. The origins  $\mathbf{R}_p$  of the primitives in (3.283) are almost always equal to  $\mathbf{R}_A$ . Different origins for the primitives in a contraction are used only with Gaussian lobe basis sets. In these basis sets one approximates  $s, p, d, \dots$  functions as combinations of spherical  $1s$  Gaussians (lobes) placed appropriately in space. For example, a  $2p$  Gaussian orbital can be approximated as closely as desired

by two 1s Gaussian lobes of opposite sign placed an infinitesimal distance apart. We will not be concerned with Gaussian lobes here.

A common way of determining contractions is from the results of atomic SCF calculations. In these atomic calculations one uses a relatively large basis of uncontracted Gaussians, optimizes all exponents, and determines the SCF coefficients of each of the derived atomic orbitals. The optimized exponents and SCF coefficients can then be used to derive suitable contraction exponents and contraction coefficients for a smaller basis set to be used in subsequent molecular calculations. Let us first illustrate this with s-type basis functions for hydrogen. Huzinaga<sup>8</sup> has determined coefficients and exponents of Gaussian expansions that minimize the energy of a hydrogen atom. With four Gaussian functions he obtains

$$\begin{aligned}\psi_{1s} = & 0.50907g_{1s}(0.123317, \mathbf{r}) + 0.47449\tilde{g}_{1s}(0.453757, \mathbf{r}) \\ & + 0.13424g_{1s}(2.01330, \mathbf{r}) + 0.01906g_{1s}(13.3615, \mathbf{r})\end{aligned}\quad (3.287)$$

The basis set is an uncontracted basis consisting of four functions of s-type symmetry, i.e., it is a (4s) basis. A contracted basis set derived from this would use the four Gaussian functions as primitives and contract them to reduce the number of basis functions. There are a number of ways the above four primitives might be contracted. One usually uses disjoint subsets of primitives so that no primitive appears in more than one basis function. From evidence on molecular calculations, it appears that a useful contraction scheme is one which leaves the most diffuse primitive uncontracted and contracts the remaining three primitives into one basis function, with contraction coefficients just equal to the above coefficients (SCF coefficients in the general case). That is,

$$\phi_1(\mathbf{r}) = g_{1s}(0.123317, \mathbf{r}) \quad (3.288)$$

$$\begin{aligned}\phi_2(\mathbf{r}) = & N[0.47449g_{1s}(0.453757, \mathbf{r}) + 0.13424g_{1s}(2.01330, \mathbf{r}) \\ & + 0.01906g_{1s}(13.3615, \mathbf{r})] \\ = & 0.817238g_{1s}(0.453757, \mathbf{r}) + 0.231208g_{1s}(2.01330, \mathbf{r}) \\ & + 0.032828g_{1s}(13.3615, \mathbf{r})\end{aligned}\quad (3.289)$$

In the last equation, the contraction coefficients have been properly renormalized. This scheme leads to a contracted basis set of two s-type functions, i.e., a [2s] contracted basis set, coming from a (4s) uncontracted basis set. This defines a (4s)/[2s] contraction.

Huzinaga also determined relatively large uncontracted Gaussian (9s5p) basis sets, with optimized exponents, for the first-row atoms Li to Ne. Dunning<sup>9</sup> has suggested useful contractions of these. As an example of the procedure, consider a [3s2p] contracted basis for the oxygen atom. We are going to contract the nine primitives of s type into three basis functions. On inspecting the atomic SCF calculation we find that one of the nine primitives

l an infinitesimal distance lobes here.

from the results of atomic one uses a relatively large exponents, and determines ic orbitals. The optimized to derive suitable contrac- smaller basis set to be used t illustrate this with *s*-type determined coefficients and : the energy of a hydrogen

$$9g_{1s}(0.453757, \mathbf{r}) \\ 106g_{1s}(13.3615, \mathbf{r}) \quad (3.287)$$

; of four functions of *s*-type basis set derived from this itives and contract them to a number of ways the above illy uses disjoint subsets of re than one basis function. pears that a useful contrac- primitive uncontracted and ne basis function, with con- ficients (SCF coefficients in

$$(3.288)$$

$$13424g_{1s}(2.01330, \mathbf{r})$$

ts have been properly renor- is set of two *s*-type functions, a (4s) uncontracted basis set.

uncontracted Gaussian (9s5p) e first-row atoms Li to Ne. of these. As an example of the for the oxygen atom. We are into three basis functions. On that one of the nine primitives

contributes strongly to both the 1s and 2s orbitals of the oxygen atom; this function is left uncontracted.

$$\phi_1(\mathbf{r}) = g_{1s}(9.5322, \mathbf{r}) \quad (3.290)$$

The two primitives, which are most diffuse, make a negligible contribution to the 1s atomic orbital but are the chief contributors to the 2s atomic orbital. They are contracted to give the second basis function,

$$\begin{aligned} \phi_2(\mathbf{r}) &= N[0.59566g_{1s}(0.9398, \mathbf{r}) + 0.52576g_{1s}(0.2846, \mathbf{r})] \\ &= 0.563459g_{1s}(0.9398, \mathbf{r}) + 0.497338g_{1s}(0.2846, \mathbf{r}) \end{aligned} \quad (3.291)$$

where 0.59566 and 0.52576 are the coefficients of these primitives in the 2s atomic orbital of the atomic SCF calculation. The last basis function consist of the remainder of the nine primitives,

$$\begin{aligned} \phi_3(\mathbf{r}) &= N[0.14017g_{1s}(3.4136, \mathbf{r}) + 0.35555g_{1s}(27.1836, \mathbf{r}) \\ &\quad + 0.14389g_{1s}(81.1696, \mathbf{r}) + 0.04287g_{1s}(273.188, \mathbf{r}) \\ &\quad + 0.00897g_{1s}(1175.82, \mathbf{r}) + 0.00118g_{1s}(7816.54, \mathbf{r})] \\ &= 0.241205g_{1s}(3.4136, \mathbf{r}) + 0.611832g_{1s}(27.1836, \mathbf{r}) \\ &\quad + 0.247606g_{1s}(81.1696, \mathbf{r}) + 0.073771g_{1s}(273.188, \mathbf{r}) \\ &\quad + 0.015436g_{1s}(1175.82, \mathbf{r}) + 0.002031g_{1s}(7816.54, \mathbf{r}) \end{aligned} \quad (3.292)$$

where 0.14017, 0.35555, etc, are the coefficients of these primitives in the 1s atomic orbital of the atomic SCF calculation.

In a similar way, the five primitives of *p*-type symmetry are contracted to two basis functions. Here the most diffuse *p*-function is left uncontracted,

$$\phi_1(\mathbf{r}) = g_{2p}(0.2137, \mathbf{r}) \quad (3.293)$$

and the remaining four primitives are contracted using the SCF coefficients of the 2*p* atomic orbital

$$\begin{aligned} \phi_2(\mathbf{r}) &= N[0.49376g_{2p}(0.7171, \mathbf{r}) + 0.31066g_{2p}(2.3051, \mathbf{r}) \\ &\quad + 0.09774g_{2p}(7.9040, \mathbf{r}) + 0.01541g_{2p}(35.1832, \mathbf{r})] \\ &= 0.627375g_{2p}(0.71706, \mathbf{r}) + 0.394727g_{2p}(2.30512, \mathbf{r}) \\ &\quad + 0.124189g_{2p}(7.90403, \mathbf{r}) + 0.019580g_{2p}(35.1835, \mathbf{r}) \end{aligned} \quad (3.294)$$

This (9s5p)/[3s2p] contraction reduces the number of basis functions from 24 to 9. Remember the  $p_x$ ,  $p_y$ , and  $p_z$  are included for each *p* orbital exponent. A calculation with either basis set would give almost identical results in a calculation on the oxygen atom, however. The loss of variational flexibility in molecular calculations is not extreme either. For example, a calculation on the water molecule using the fully uncontracted (9s5p/4s)<sup>10</sup> basis set gives an energy of -76.0133, whereas the [3s2p/2s] contracted basis gives an energy of -76.0080, only 0.007% above the much larger calculation. Because the cost of an SCF calculation increases with the fourth

power of the number of basis functions, the reduction from 32 functions to 13 functions is impressive.

### 3.6.2 Minimal Basis Sets: STO-3G

A minimal basis set is a relatively inexpensive one, which can be used for calculations on quite large molecules. It is minimal in the sense of having the least number of functions per atom required to describe the occupied atomic orbitals of that atom. This is not quite accurate, since one usually considers 1s, 2s and 2p, i.e., five functions, to constitute a minimal basis set for Li and Be, for example, even though the 2p orbital is not occupied in these atoms. The 2sp (2s and 2p), 3sp, 4sp, 3d, . . . , etc. shells are considered together. The minimal basis set thus consists of 1 function for H and He, 5 functions for Li to Ne, 9 functions for Na to Ar, 13 functions for K and Ca, 18 functions for Sc to Kr, . . . , etc. Because the minimal basis set is so small, it is not one which can lead to quantitatively accurate results. It does, however, contain the essentials of chemical bonding and many useful qualitative results can be obtained with it.

Because of the small number of functions in a minimal basis set, it is particularly important that these functions be of near optimum form. This immediately rules out a single Gaussian function. One would prefer to use Slater functions or functions that closely resemble the known shape of atomic orbitals. A significant advance in minimal basis calculations came with the development of computer programs like "Gaussian 70," which could reproduce the results of minimal basis Slater orbital calculations using contracted Gaussian functions. The STO-LG method uses a contraction of  $L$  primitive Gaussians for each basis function, where the contraction coefficients and exponents are chosen so that the basis functions approximate Slater functions. We have already discussed the 1s STO-3G basis set in Subsection 3.5.1.

The calculations in this book are restricted to a small number of molecules, all of which include only the first row atoms up to fluorine. Although the STO-LG method has been extended to second row atoms, we will only consider its formulation, and the formulation of the other basis sets which follow, for first row atoms and, in particular, for H, C, N, O, and F. We are therefore interested in the expansion of the 1s, 2s, and 2p Slater functions in a set of primitive Gaussians

$$\phi_{1s}^{\text{CGF}}(\zeta = 1.0) = \sum_{i=1}^L d_{i,1s} g_{1s}(\alpha_{i,1s}) \quad (3.295)$$

$$\phi_{2s}^{\text{CGF}}(\zeta = 1.0) = \sum_{i=1}^L d_{i,2s} g_{1s}(\alpha_{i,2sp}) \quad (3.296)$$

$$\phi_{2p}^{\text{CGF}}(\zeta = 1.0) = \sum_{i=1}^L d_{i,2p} g_{2p}(\alpha_{i,2sp}) \quad (3.297)$$

where the . . . by a least-

and

One of the . . . is the shar . . . exponents i . . . 2p fits are above. The same expo . . . radial part That is, all radial integ . . . grouping of siderable eff . . . uses contrac . . . however, the determined properties t . . . calculation, basis calcula . . . coefficients o . . . contraction

Once th . . .  $\zeta = 1.0$  (Tab . . . exponents ca . . . by  $\zeta^2$ . It rem . . . electronic st . . . atom" expo . . . each calculat

Table 3.7 STO-3G basis functions for 1s, 2s, and 2p

$\alpha_{1s}$	
0.109818	0
0.405771	0
2.22766	0

where the contraction coefficients ( $d$ 's) and exponents ( $\alpha$ 's) are to be obtained by a least-squares fit which minimizes the integrals

$$\int d\mathbf{r} [\phi_{1s}^{\text{SF}}(\mathbf{r}) - \phi_{1s}^{\text{CGF}}(\mathbf{r})]^2$$

and

$$\int d\mathbf{r} [\phi_{2s}^{\text{SF}}(\mathbf{r}) - \phi_{2s}^{\text{CGF}}(\mathbf{r})]^2 + \int d\mathbf{r} [\phi_{2p}^{\text{SF}}(\mathbf{r}) - \phi_{2p}^{\text{CGF}}(\mathbf{r})]^2$$

One of the unique aspects of the STO-LG method and the fitting procedure is the sharing of contraction exponents in  $2sp$ ,  $3sp$ , ... shells. Thus the exponents in (3.296) and (3.297) are constrained to be identical and the  $2s$  and  $2p$  fits are performed simultaneously as indicated by the second integral above. The reason for this constraint is that if  $2s$  and  $2p$  functions have the same exponents, then they have the same radial behavior, and during the radial part of the integral evaluation they can be treated as one function. That is, all integrals involving any  $sp$  shell are treated together and one radial integration is sufficient for up to  $256 \equiv 4^4$  separate integrals. This grouping of basis functions by shells with shared exponents leads to considerable efficiency in integral evaluation. The general STO-LG procedure uses contraction lengths up to  $L = 6$ . The longer the length of the contraction, however, the more time is spent in integral evaluation. It has been empirically determined that a contraction of length 3 is sufficient to lead to calculated properties that reproduce essentially all the valence features of a Slater calculation, and STO-3G has become the *de facto* standard for minimal basis calculations. Table 3.7 gives the STO-3G contraction exponents and coefficients of Eqs. (3.295) and (3.297). In the general notation, the STO-3G contraction is  $(6s3p/3s)/[2s1p/1s]$ .

Once the least-squares fits to Slater functions with orbital exponents  $\zeta = 1.0$  (Table 3.7) are available, fits to Slater functions with other orbital exponents can be obtained by simply multiplying the  $\alpha$ 's in (3.295) to (3.297) by  $\zeta^2$ . It remains to be determined what Slater orbital exponents  $\zeta$  to use in electronic structure calculations. Two possibilities might be to use "best atom" exponents ( $\zeta = 1.0$  for H, for example) or to optimize exponents in each calculation. The "best atom" exponents might be rather a poor choice

**Table 3.7** STO-3G contraction exponents and coefficients for  $1s$ ,  $2s$ , and  $2p$  basis functions

$\alpha_{1s}$	$d_{1s}$	$\alpha_{2sp}$	$d_{2s}$	$d_{2p}$
0.109818	0.444635	0.0751386	0.700115	0.391957
0.405771	0.535328	0.231031	0.399513	0.607684
2.22766	0.154329	0.994203	-0.0999672	0.155916

**Table 3.8 Standard STO-3G exponents**

Atom	$\zeta_{1s}$	$\zeta_{2sp}$
H	1.24	—
Li	2.69	0.75
Be	3.68	1.10
B	4.68	1.45
C	5.67	1.72
N	6.67	1.95
O	7.66	2.25
F	8.65	2.55

for molecular environments, and optimization of nonlinear exponents is not practical for large molecules, where the dimension of the space to be searched is very large. A compromise is to use a set of standard exponents which are the average values of exponents optimized for a set of small molecules. The recommended STO-3G exponents are shown in Table 3.8.

The STO-LG basis is not the only possible minimal basis of course. Stewart,<sup>11</sup> for example, has determined fits of contracted Gaussian functions to individual Slater functions, without the constraint of sharing exponents in a shell. Rather than use Slater functions or fits to Slater functions, a reasonable choice is contracted basis functions which closely approximate the individually determined Hartree-Fock atomic orbitals of the atom. Calculations suggest, however, that Slater functions with near optimum exponents are better than these Hartree-Fock atomic orbitals for a minimal basis; orbitals in molecules may be rather different than those in the constituent atoms.

### 3.6.3 Double Zeta Basis Sets: 4-31G

A minimal basis set has rather limited variational flexibility particularly if exponents are not optimized. The first step in improving upon the minimal basis set involves using two functions for each of the minimal basis functions—a double zeta basis set. The best orbital exponents of the two functions are commonly slightly above and slightly below the optimal exponent of the minimal basis function. This allows effective expansion or “contraction” of the basis functions by variation of linear parameters rather than nonlinear exponents. The SCF procedure will weight either the coefficient of the dense or diffuse component according to whether the molecular environment requires the effective orbital to be expanded or “contracted.” In addition, an extra degree of anisotropy is allowed relative to an STO-3G basis since, for example, *p* orbitals in different directions can have effectively different sizes.

The 4-3 valence func inner shell o shells contri slightly from has some eff valence ioniz most other c consists of 2 for Na to Ar,

The outer function  $\phi'_{1s}$  is numerical diff the above bas and (3.289). T form but is de mizing the en coefficients an basis function function) and c shell functions course, does no

For the ato

As in the S computational e coefficients  $d_{1s}$ ,  $\alpha'_{2sp}$ , and  $\alpha''_{2sp}$  v calculation reac

The 4-31G basis set is not exactly a double zeta basis since only the valence functions are doubled and a single function is still used for each inner shell orbital. It may be termed a split valence shell basis set. The inner shells contribute little to most chemical properties and usually vary only slightly from molecule to molecule. Not splitting the inner shell functions has some effect on the total energy, but little effect on dipole moments, valence ionization potentials, charge densities, dissociation energies, and most other calculated quantities of chemical interest. The 4-31G basis thus consists of 2 functions for H and He, 9 functions for Li to Ne, 13 functions for Na to Ar, . . . , etc. For hydrogen the contractions are

$$\phi'_{1s}(\mathbf{r}) = \sum_{i=1}^3 d'_{i,1s} g_{1s}(\alpha'_{i,1s}, \mathbf{r}) \quad (3.298)$$

$$\phi''_{1s}(\mathbf{r}) = g_{1s}(\alpha''_{1s}, \mathbf{r}) \quad (3.299)$$

The outer hydrogen function  $\phi''_{1s}$  is uncontracted and the inner hydrogen function  $\phi'_{1s}$  is a contraction of three primitive Gaussians. Apart from small numerical differences in deriving the contraction coefficients and exponents, the above basis functions are identical to the (4s)/[2s] functions of (3.288) and (3.289). That is, the 4-31G basis is not fit to any particular functional form but is derived by choosing the form of the contraction and then minimizing the energy of an atomic calculation by varying the contraction coefficients and exponents. The 4-31G acronym implies that the valence basis functions are contractions of three primitive Gaussians (the inner function) and one primitive Gaussian (the outer function), whereas the inner shell functions are contractions of four primitive Gaussians. Hydrogen, of course, does not have inner shells.

For the atoms Li to F, the contractions are

$$\phi_{1s}(\mathbf{r}) = \sum_{i=1}^4 d_{i,1s} g_{1s}(\alpha_{i,1s}, \mathbf{r}) \quad (3.300)$$

$$\phi'_{2s}(\mathbf{r}) = \sum_{i=1}^3 d'_{i,2s} g_{1s}(\alpha'_{i,2sp}, \mathbf{r}) \quad (3.301)$$

$$\phi''_{2s}(\mathbf{r}) = g_{1s}(\alpha''_{2sp}, \mathbf{r}) \quad (3.302)$$

$$\phi'_{2p}(\mathbf{r}) = \sum_{i=1}^3 d'_{i,2p} g_{2p}(\alpha'_{i,2sp}, \mathbf{r}) \quad (3.303)$$

$$\phi''_{2p}(\mathbf{r}) = g_{2p}(\alpha''_{2sp}, \mathbf{r}) \quad (3.304)$$

As in the STO-3G basis, the 2s and 2p functions share exponents for computational efficiency. Given the above functional forms, the contraction coefficients  $d_{1s}$ ,  $d'_{2s}$ ,  $d''_{2s}$ ,  $d'_{2p}$ , and  $d''_{2p}$ , and the contraction exponents  $\alpha_{1s}$ ,  $\alpha'_{2sp}$ , and  $\alpha''_{2sp}$  were explicitly varied until the energy of an atomic SCF calculation reached a minimum. Unlike the STO-3G basis, which was

**Table 3.9 Standard 4-31G valence shell scale factors**

Atom	$\zeta'$	$\zeta''$
H	1.20	1.15
C	1.00	1.04
N	0.99	0.98
O	0.99	0.98
F	1.00	1.00

obtained by a least-squares fit to known functions, or a general contraction scheme based on contraction of previously determined uncontracted atomic calculations, the 4-31G basis sets were determined by choosing the specific form (3.300) to (3.304) for the contractions and then optimizing all contraction parameters. That is, the basis set was obtained by contraction first, then optimization, as opposed to optimization first, then contraction. In our general notation, the 4-31G contraction is written as  $(8s4p/4s)/[3s2p/2s]$ . The basis consists of inner shell functions, inner valence functions, and outer valence functions. These are contractions of 4, 3, and 1 primitive functions, respectively.

Since the basis set is obtained from atomic calculations, it is still desirable to scale exponents for the molecular environment. This is accomplished by defining an inner valence scale factor  $\zeta'$  and an outer valence scale factor  $\zeta''$  and multiplying the corresponding inner and outer  $\alpha$ 's by the square of these factors. Only the valence shells are scaled. Table 3.9 gives a set of standard 4-31G scale factors. Only those for H differ significantly from unity, although the outer carbon functions are somewhat denser than in the atom.

**Exercise 3.30** A 4-31G basis for He has not been officially defined. Huzinaga,<sup>8</sup> however, in an SCF calculation on the He atom using four uncontracted 1s Gaussians, found the coefficients and optimum exponents of the normalized 1s orbital of He to be

$\alpha_\mu$	$C_{\mu i}$
0.298073	0.51380
1.242567	0.46954
5.782948	0.15457
38.47497	0.02373

Use the expression for overlaps given in Appendix A to derive the contraction parameters for a 4-31G He basis set.

### 3.6.4 Polarized E

The next step in im zeta, etc. If one goes angular quantum t limit of a large nu that the equilibriu next step beyond d i.e., adding  $d$ -type fu to H. To see why t hydrogen atom. The just the 1s orbital. If however, the electro and the charge dist polarized. The low original 1s orbital ar to be a hybridized c similar, but nonunifo ment. By adding pola for H we directly acc which are not occup functions for the ato resemble the 4-31G b atoms (\*) or  $d$ -type fu added to hydrogen ( polarization functions polarization functions STO-3G, 4-31G, 6-31C

The 6-31G\* and 6 function to a 6-31G ba to those of the 4-31G b Li to F) become a con The 6-31G optimizatio valence functions are n similar. The 6-31G and valence properties altho of the improvement in t

The  $d$ -type function basis are a single set c putational convenience  $3d_{zz}$ ,  $3d_{xy}$ ,  $3d_{yz}$ , and  $3d_{zx}$  combinations of the us  $3d_{z^2}$  and a 3s function ( $x^2$  polarization functions to

### 3.6.4 Polarized Basis Sets: 6-31G\* and 6-31G\*\*

The next step in improving a basis set could be to go to triple zeta, quadruple zeta, etc. If one goes in this direction rather than adding functions of higher angular quantum number, the basis set would not be well balanced. In the limit of a large number of only *s* and *p* functions, one finds, for example, that the equilibrium geometry of ammonia actually becomes planar. The next step beyond double zeta usually involves adding *polarization functions*, i.e., adding *d*-type functions to the first row atoms Li-F and *p*-type functions to H. To see why these are called polarization functions, consider the hydrogen atom. The exact wave function for an isolated hydrogen atom is just the 1s orbital. If the hydrogen atom is placed in a uniform electric field, however, the electron cloud is attracted to the direction of the electric field, and the charge distribution about the nucleus becomes asymmetric. It is polarized. The lowest order solution to this problem is a mixture of the original 1s orbital and a *p*-type function, i.e., the solution can be considered to be a hybridized orbital. A hydrogen atom in a molecule experiences a similar, but nonuniform, electric field arising from its nonspherical environment. By adding polarization functions, i.e., *p*-type functions, to a basis set for H we directly accommodate this effect. In a similar way, *d*-type functions, which are not occupied in first row atoms, play the role of polarization functions for the atoms Li to F. The 6-31G\* and 6-31G\*\* basis sets closely resemble the 4-31G basis set with *d*-type basis functions added to the heavy atoms (\*) or *d*-type functions added to the heavy atoms, and *p*-type functions added to hydrogen (\*\*). It has been empirically determined that adding polarization functions to the heavy atoms is more important than adding polarization functions to hydrogen. The hierarchy of our basis sets is thus STO-3G, 4-31G, 6-31G\*, and 6-31G\*\*.

The 6-31G\* and 6-31G\*\* basis sets are formed by adding polarization function to a 6-31G basis. The form of the 6-31G contractions are identical to those of the 4-31G basis, except that the inner shell functions (1s only, for Li to F) become a contraction of six primitive Gaussians rather than four. The 6-31G optimization was performed from the beginning and so the valence functions are not identical to those of the 4-31G basis, but are very similar. The 6-31G and 4-31G basis sets give almost identical results for valence properties although the 6-31G basis gives lower energies, because of the improvement in the inner shell.

The *d*-type functions that are added to a 6-31G basis to form a 6-31G\* basis are a single set of uncontracted 3*d* primitive Gaussians. For computational convenience there are "six 3*d* functions" per atom— $3d_{xx}$ ,  $3d_{yy}$ ,  $3d_{zz}$ ,  $3d_{xy}$ ,  $3d_{yz}$ , and  $3d_{zx}$ . These six, the Cartesian Gaussians, are linear combinations of the usual five 3*d* functions— $3d_{xy}$ ,  $3d_{x^2-y^2}$ ,  $3d_{yz}$ ,  $3d_{zx}$ , and  $3d_{z^2}$  and a 3s function ( $x^2 + y^2 + z^2$ ). The 6-31G\* basis, in addition to adding polarization functions to a 6-31G basis, thus includes one more function of

*s*-type symmetry. The contraction is thus  $(11s4p1d/4s)/[4s2p1d/2s]$  and the basis set includes 2 functions for H and 15 functions for Li to F. A standard Gaussian exponent for the six 3*d* functions of  $\alpha = 0.8$  has been suggested for C, N, O, and F.

The 6-31G\*\* basis differs from the 6-31G\* basis by the addition of one set of uncontracted *p*-type Gaussian primitives for each H. A standard Gaussian exponent of  $\alpha = 1.1$  has been suggested for these functions. The 6-31G\*\* contraction is thus  $(11s4p1d/4s1p)[4s2p1d/2s1p]$  and each hydrogen now includes five basis functions.

---

**Exercise 3.31** Determine the total number of basis functions for STO-3G, 4-31G, 6-31G\*, and 6-31G\*\* calculations on benzene.

---

Calculations at the 6-31G\* and 6-31G\*\* level provide, in many cases, quantitative results considerably superior to those at the lower STO-3G and 4-31G levels. Even these basis sets, however, have deficiencies that can only be remedied by going to triple zeta or quadruple zeta, adding more than one set of polarization functions, adding *f*-type functions to heavy atoms and *d*-type functions to hydrogen, improving the basis function description of the inner shell electrons, etc. As technology improves it will be possible to use more and more accurate basis sets.

### 3.7 SOME ILLUSTRATIVE CLOSED-SHELL CALCULATIONS

In this section, we illustrate results that are characteristic of Hartree-Fock calculations on the ground state of closed-shell molecules. Now that we have discussed polyatomic basis sets and the closed-shell restricted Hartree-Fock procedure, we are in a position to appreciate the results and the methodology of sample SCF calculations. The results of an extremely large number of SCF calculations are now available in the literature; we make no attempt to review these calculations. Instead, we apply a well-defined hierarchy of basis sets to a small set of "typical" molecules, and use these calculations to illustrate the order of accuracy expected in the general SCF calculation. By restricting our calculations to a few well-defined basis sets and a small set of molecules, we will be able to apply the various methods of later chapters, which go beyond the Hartree-Fock approximation, to the same collection of basis sets and molecules. In this way, we hope to give a more systematic illustration of the results obtained from the many computational methods of quantum chemistry than would be possible by simply reviewing selected results available in the literature. Thus our purpose here, in addition to illustrating SCF results themselves, is to display Hartree-Fock values of a small number of calculated quantities, for comparison with

**Table 3.10 Standard calculations**

Molecule	Bond length (Å)
H <sub>2</sub>	1.40
CO	2.11
N <sub>2</sub>	2.07
CH <sub>4</sub>	2.05
NH <sub>3</sub>	1.91
H <sub>2</sub> O	1.80
FH	1.73

better values obtained in results of this section. These errors are corrected.

The molecules we have used and the ten-electron systems at which all calculations are shown in Table 3.10 are shown in Table 3.10. They are always identical to values determined by the semi-empirical method. The semi-empirical calculations illustrate the total weak interactions. They do not illustrate the total weak interactions that can be derived from calculations in the next section. For the most part, however, they are performed on the molecules.

#### 3.7.1 Total Energies

Perhaps the primary quantity of interest in quantum mechanics is the total energy. The total energy is given by the quantum mechanical calculation. In the SCF approximation, the total energy is given by the "better" basis set, the more and more complete basis set. This limit can be approached by the variational principle. By the variational principle, the exact energy, which is the exact solution to the Schrödinger equation, is given by the Hartree-Fock approximation. He, Be, etc., proper account of the Hartree-Fock corrections when