

# Molecular Electronic Structure

## 15.1 Ab Initio, Density-Functional, Semiempirical, and Molecular-Mechanics Methods

The electronic wave function of a polyatomic molecule depends on several parameters—the bond distances, bond angles, and dihedral angles of rotation about single bonds (these angles define the molecular conformation). A full theoretical treatment of a polyatomic molecule involves calculation of the electronic wave function for a range of each of these parameters. The equilibrium bond distances, bond angles, and dihedral angles are then found as those values that minimize the electronic energy including nuclear repulsion.

The four main approaches to calculating molecular properties are *ab initio* methods, semiempirical methods, the density-functional method, and the molecular-mechanics method.

**Semiempirical** molecular quantum-mechanical methods use a simpler Hamiltonian than the correct molecular Hamiltonian and use parameters whose values are adjusted to fit experimental data or the results of *ab initio* calculations. An example is the Hückel MO treatment of conjugated hydrocarbons (Section 17.2), which uses a one-electron Hamiltonian and takes the bond integrals as adjustable parameters rather than quantities to be calculated theoretically. In contrast, an ***ab initio*** (or **first principles**) calculation uses the correct Hamiltonian and does not use experimental data other than the values of the fundamental physical constants. A Hartree–Fock SCF calculation seeks the antisymmetrized product  $\Phi$  of one-electron functions that minimizes  $\int \Phi^* \hat{H} \Phi d\tau$ , where  $\hat{H}$  is the true Hamiltonian and is thus an *ab initio* calculation. (*Ab initio* is Latin for “from the beginning” and indicates a calculation based on fundamental principles.) The term *ab initio* should not be interpreted to mean “100% correct.” An *ab initio* SCF MO calculation uses the approximation of taking  $\psi$  as an antisymmetrized product of one-electron spin-orbitals and uses a finite (and hence incomplete) basis set.

The **density-functional method** (Section 16.5) does not attempt to calculate the molecular wave function but calculates the molecular electron probability density  $\rho$  and calculates the molecular electronic energy from  $\rho$ .

The **molecular-mechanics method** (Section 17.5) is not a quantum-mechanical method and does not use a molecular Hamiltonian operator or wave function. Instead, it views the molecule as a collection of atoms held together by bonds and expresses the molecular energy in terms of force constants for bond bending, stretching, and torsion and other parameters.

This chapter discusses general principles of electronic structure calculations for polyatomic molecules and discusses the Hartree–Fock calculation method. As noted in Section 11.3, the Hartree–Fock method does not take account of electron correlation.

Ab initio and density-functional methods that take account of electron correlation are discussed in Chapter 16. Chapter 17 discusses semiempirical methods and the molecular-mechanics method.

A free database of references to molecular ab initio and density-functional calculations is the Quantum Chemistry Literature Database at [qcldb2.ims.ac.jp](http://qcldb2.ims.ac.jp).

The National Institute of Standards and Technology's Computational Chemistry Comparison and Benchmark Database (**CCCBDB**) at [cccbdb.nist.gov](http://cccbdb.nist.gov) tabulates the results of ab initio, density functional, and semiempirical calculations with various basis sets (Section 15.4) on 1420 molecules and includes gas-phase experimental data for the tabulated species. Calculated properties listed include energy, geometry, dipole moment, polarizability, harmonic vibrational frequencies, moments of inertia, barriers to internal rotation, gas-phase thermodynamic properties, and atomic charges (Section 15.7). The program Spartan (Section 15.14) gives access to part or all (depending on the type of license the user has) of a database of ab initio and density functional results for 140000 molecules, for a small number of basis sets.

## 15.2 Electronic Terms of Polyatomic Molecules

For polyatomic molecules, the operator  $\hat{S}^2$  for the square of the total electronic spin angular momentum commutes with the electronic Hamiltonian operator, and, as for diatomic molecules, the electronic terms of polyatomic molecules are classified as singlets, doublets, triplets, and so on, according to the value of  $2S + 1$ . (The commutation of  $\hat{S}^2$  and  $\hat{H}_{el}$  holds provided spin-orbit interaction is omitted from the Hamiltonian. For molecules containing heavy atoms, spin-orbit interaction is considerable, and  $S$  is not a good quantum number.)

For linear polyatomic molecules, the operator  $\hat{L}_z$  for the component of the total electronic orbital angular momentum along the molecular axis commutes with the electronic Hamiltonian, and the same term classifications are used as for diatomic molecules (Section 13.8), giving such possibilities as  $^1\Sigma^+$ ,  $^1\Sigma^-$ ,  $^3\Sigma^+$ ,  $^1\Pi$ , and so on. For linear polyatomic molecules with a center of symmetry, the  $g$ ,  $u$  classification is added.

For nonlinear polyatomic molecules, no orbital angular-momentum operator commutes with the electronic Hamiltonian, and the angular-momentum classification of electronic terms cannot be used. Operators that do commute with the electronic Hamiltonian are the symmetry operators  $\hat{O}_R$  of the molecule (Section 12.1), and the electronic states of polyatomic molecules are classified according to the behavior of the electronic wave function on application of these operators. Consider  $\text{H}_2\text{O}$  as an example.

In its equilibrium configuration, the water molecule belongs to group  $\mathcal{C}_{2v}$  with the symmetry operations

$$\hat{E} \quad \hat{C}_2(z) \quad \hat{\sigma}_v(xz) \quad \hat{\sigma}_v(yz) \quad (15.1)$$

The standard convention [R.S. Mulliken, *J. Chem. Phys.*, **23**, 1997 (1955); **24**, 1118 (1956)] takes the molecular plane as the  $yz$  plane (Fig. 15.1). We readily find that each of the

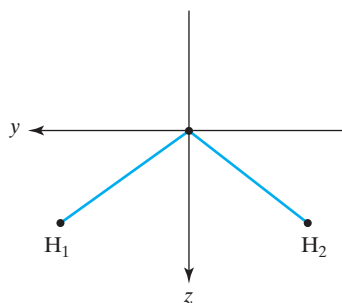


**FIGURE 15.1** Coordinate axes for the  $\text{H}_2\text{O}$  molecule. The  $x$  axis is perpendicular to the molecular plane.

symmetry operations commutes with the other three. Therefore, the electronic wave functions can be chosen as simultaneous eigenfunctions of all four symmetry operators. Since  $\hat{O}_E$  is the unit operator, we have  $\hat{O}_E\psi_{\text{el}} = \psi_{\text{el}}$ . Each of the remaining symmetry operators satisfies  $\hat{O}_R^2 = \hat{1}$ , and so each has as its eigenvalues  $+1$  and  $-1$  [Eq. (7.55)]. Therefore, each electronic wave function of  $\text{H}_2\text{O}$  is an eigenfunction of  $\hat{O}_E$  with eigenvalue  $+1$  and an eigenfunction of each of the other three symmetry operators with the eigenvalue  $+1$  or  $-1$ .

How many different combinations of these symmetry eigenvalues are there for  $\text{H}_2\text{O}$ ? At first sight, we might think there are  $(1)(2)(2)(2) = 8$  possible sets. However, certain sets can be ruled out, as we now show. Let the product of the symmetry operations  $\hat{R}$  and  $\hat{S}$  be the symmetry operation  $\hat{T}$ ;  $\hat{R}\hat{S} = \hat{T}$ . Let  $\psi_{\text{el}}$  be an eigenfunction of  $\hat{O}_R$ ,  $\hat{O}_S$ , and  $\hat{O}_T$  with eigenvalues  $r$ ,  $s$ , and  $t$ , respectively. Since the symmetry operators multiply the same way the symmetry operations do, we have

$$t\psi_{\text{el}} = \hat{O}_T\psi_{\text{el}} = \hat{O}_R(\hat{O}_S\psi_{\text{el}}) = s\hat{O}_R\psi_{\text{el}} = rs\psi_{\text{el}}$$

Dividing by  $\psi_{\text{el}}$ , we have  $rs = t$  if  $\hat{R}\hat{S} = \hat{T}$ . Hence the eigenvalues of the symmetry operators must multiply in the same way the symmetry operations do. Now consider  $\text{H}_2\text{O}$ . Let us examine the set of symmetry eigenvalues

$$\begin{array}{cccc} \hat{E} & \hat{C}_2(z) & \hat{\sigma}_v(xz) & \hat{\sigma}_v(yz) \\ 1 & -1 & -1 & -1 \end{array} \quad (15.2)$$

We find that  $\hat{C}_2(z)\hat{\sigma}_v(xz) = \hat{\sigma}_v(yz)$ . The set (15.2), however, has the product of the  $\hat{O}_{C_2}$  and  $\hat{O}_{\sigma_v(xz)}$  eigenvalues as  $(-1)(-1) = 1$ , which differs from the  $\hat{O}_{\sigma_v(yz)}$  eigenvalue in (15.2). The set (15.2) must be discarded. Of the eight possible symmetry-eigenvalue combinations, only the following four sets are found to multiply properly (Prob. 15.1):

|       | $\hat{E}$ | $\hat{C}_2(z)$ | $\hat{\sigma}_v(xz)$ | $\hat{\sigma}_v(yz)$ |
|-------|-----------|----------------|----------------------|----------------------|
| $A_1$ | 1         | 1              | 1                    | 1                    |
| $A_2$ | 1         | 1              | -1                   | -1                   |
| $B_1$ | 1         | -1             | 1                    | -1                   |
| $B_2$ | 1         | -1             | -1                   | 1                    |

(15.3)

In this table the sets have been labeled  $A_1$ ,  $A_2$ ,  $B_1$ , and  $B_2$ . The letter  $A$  or  $B$  indicates whether the symmetry eigenvalue for the highest-order  $\hat{C}_n$  or  $\hat{S}_n$  operation of the molecule [ $\hat{C}_2(z)$  for water] is  $+1$  or  $-1$ , respectively. The subscripts 1 and 2 distinguish sets having the same letter label. Each possible set of symmetry eigenvalues in (15.3) is called a **symmetry species** (or *symmetry type*). (The group-theory term is **irreducible representation**.) The symmetry species with all symmetry eigenvalues  $+1$  ( $A_1$  for  $\text{H}_2\text{O}$ ) is called the **totally symmetric species**.

Each molecular electronic term of  $\text{H}_2\text{O}$  is designated by giving the symmetry species of the electronic wave functions of the term, with the spin multiplicity  $2S + 1$  as a left superscript. For example, an electronic state of  $\text{H}_2\text{O}$  with two electrons unpaired and with the electronic wave function unchanged by all four symmetry operators belongs to a  $^3A_1$  term. (The subscript 1 is not an angular-momentum eigenvalue but is part of the symmetry-species label.)

We now consider the **orbital degeneracy** of molecular electronic terms. This is degeneracy connected with the electrons' spatial (orbital) motion, as distinguished from spin degeneracy. Thus  $^1\Pi$  and  $^3\Pi$  terms of linear molecules are orbitally degenerate, while  $^1\Sigma$  and  $^3\Sigma$  terms are orbitally nondegenerate. Consider an operator  $\hat{F}$  that

commutes with the molecular electronic Hamiltonian and that does not involve spin. We have

$$\begin{aligned}\hat{F}\hat{H}_{\text{el}}\psi_{\text{el},i} &= \hat{F}E_{\text{el}}\psi_{\text{el},i} \\ \hat{H}_{\text{el}}(\hat{F}\psi_{\text{el},i}) &= E_{\text{el}}(\hat{F}\psi_{\text{el},i})\end{aligned}\quad (15.4)$$

Thus  $\hat{F}\psi_{\text{el},i}$  is an eigenfunction of  $\hat{H}_{\text{el}}$  with eigenvalue  $E_{\text{el}}$ . Let the orbital degeneracy of the term to which  $\psi_{\text{el},i}$  belongs be  $n$ . It follows from (15.4) that  $\hat{F}\psi_{\text{el},i}$  must be some linear combination of the  $n$  orbitally degenerate wave functions of the term that have the same value of  $S_z$  as  $\psi_{\text{el},i}$ :

$$\hat{F}\psi_{\text{el},i} = \sum_{j=1}^n c_{ij,F} \psi_{\text{el},j} \quad (15.5)$$

where the  $c$ 's are certain constants. (These arguments were previously given in Section 7.3.) For a level that is not orbitally degenerate ( $n = 1$ ), Eq. (15.5) reduces to

$$\hat{F}\psi_{\text{el},i} = c_F \psi_{\text{el},i} \quad (15.6)$$

and here  $\psi_{\text{el},i}$  must be an eigenfunction of  $\hat{F}$ . For  $n > 1$ , we can choose  $n$  linear combinations of the  $\psi_{\text{el},j}$  functions that are eigenfunctions of  $\hat{F}$ , but there is no necessity for the eigenfunctions of a degenerate level to be eigenfunctions of  $\hat{F}$ . If we have two operators  $\hat{F}$  and  $\hat{G}$  that commute with  $\hat{H}_{\text{el}}$  but not with each other, we can pick the linear combinations to be eigenfunctions of  $\hat{F}$  or of  $\hat{G}$ , but not in general of both  $\hat{F}$  and  $\hat{G}$  simultaneously.

Now consider symmetry operators. For  $\text{H}_2\text{O}$  all the symmetry operators commute among themselves, and so each electronic wave function is simultaneously an eigenfunction of all the symmetry operators. If all the electronic wave functions were orbitally nondegenerate, then it would automatically follow from (15.6) that they were simultaneously eigenfunctions of all the symmetry operators. We therefore suspect (but have not proved) that the electronic wave functions of  $\text{H}_2\text{O}$  are all orbitally nondegenerate. This statement is in fact correct. The letters  $A$  and  $B$  designate symmetry species of orbitally nondegenerate electronic terms. For any molecule all of whose symmetry operators commute with one another, the electronic wave functions are each simultaneous eigenfunctions of all the symmetry operators, and the only symmetry species are nondegenerate  $A$  and  $B$  species.

For some point groups, the symmetry operators do not all commute. An example is  $\mathcal{O}_h$  (Fig. 12.7). When the symmetry operators do not all commute, some of the electronic terms are orbitally degenerate. A symmetry operator applied to an electronic wave function of an orbitally degenerate term converts it to a linear combination of the wave functions of the term [Eq. (15.5)]. The effects of the symmetry operator  $\hat{O}_R$  on the wave functions of an  $n$ -fold orbitally degenerate electronic term are specified by the  $n^2$  numbers

$$c_{ij,R}, \quad i = 1, \dots, n, \quad j = 1, \dots, n \quad (15.7)$$

The  $n^2$  numbers (15.7) when arranged in an  $n \times n$  square array form a matrix (Section 7.10). If there are  $h$  symmetry operations in the molecular point group, then the  $h$  matrices of coefficients in (15.5) constitute the **symmetry species** of the degenerate-term wave functions in (15.5). The following letter labels are used for the symmetry species, according to the orbital degeneracy  $n$ :

|        |        |     |     |     |     |
|--------|--------|-----|-----|-----|-----|
| $n$    | 1      | 2   | 3   | 4   | 5   |
| Letter | $A, B$ | $E$ | $T$ | $G$ | $H$ |

(15.8)

Numerical subscripts distinguish different symmetry species having the same letter designation. For molecules with a center of symmetry, a  $g$  or  $u$  subscript is added, depending

on whether the wave function has the eigenvalue  $+1$  or  $-1$  for inversion of all electronic spatial coordinates. The possible symmetry species can be found in a systematic way using group theory (see *Schonland*). As an example, group theory shows the possible symmetry species of a  $\mathcal{D}_{6h}$  molecule to be

$$\begin{array}{cccccc} A_{1g} & A_{2g} & B_{1g} & B_{2g} & E_{1g} & E_{2g} \\ A_{1u} & A_{2u} & B_{1u} & B_{2u} & E_{1u} & E_{2u} \end{array} \quad (15.9)$$

We shall not give the numbers  $c_{ij,R}$  that specify these symmetry species, except to note that  $A_{1g}$  is the totally symmetric symmetry species, with all  $c$ 's equal to 1.

It is an empirical fact that for most molecules in their electronic ground states the electronic wave function belongs to the (nondegenerate) totally symmetric species. Also, the electronic spins are usually all paired in the ground state, and the ground state is a singlet. For  $\text{H}_2\text{O}$  the ground electronic state is  $^1A_1$ ; for benzene it is  $^1A_{1g}$ .

We have been using the equilibrium-geometry point groups of  $\text{H}_2\text{O}$  and  $\text{C}_6\text{H}_6$ . For nonequilibrium nuclear configurations, the symmetry is in general less than that of  $\mathcal{C}_{2v}$  or  $\mathcal{D}_{6h}$ . For reasonably small departures from equilibrium, however, the symmetry behavior should be given to a good approximation by the symmetry species of the equilibrium-geometry point group. Excited states sometimes differ in their equilibrium point group from the ground electronic state. For example, the point group of  $\text{NH}_3$  is  $\mathcal{D}_{3h}$  for several excited electronic states. (For properties of molecular electronic states, see G. Herzberg, *Electronic Spectra of Polyatomic Molecules*, Van Nostrand, 1966, Appendix VI.)

Corresponding to a given molecular electronic term, there are in general several electronic states. The interactions between electronic spin and electronic orbital motion and between electronic and nuclear motions split the energies of these states. These splittings are usually small.

## 15.3 The SCF MO Treatment of Polyatomic Molecules

The purely electronic nonrelativistic Hamiltonian for a polyatomic molecule is (in atomic units)

$$\hat{H}_{\text{el}} = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_i \sum_{\alpha} \frac{Z_{\alpha}}{r_{i\alpha}} + \sum_i \sum_{j>i} \frac{1}{r_{ij}} \quad (15.10)$$

If interelectronic repulsions are neglected, the zeroth-order wave function is the product of one-electron spatial functions (molecular orbitals). Allowance for electron spin and antisymmetry gives a zeroth-order wave function that is an antisymmetrized product of molecular spin-orbitals, each spin-orbital being a product of a spatial MO and a spin function. The best possible variation function that has the form of an antisymmetrized product of spin-orbitals is the Hartree–Fock SCF function. [Improvements beyond the Hartree–Fock stage require some method (Chapter 16) to allow for electron correlation.] The MOs are usually expressed as linear combinations of basis functions, the coefficients being found by solution of the Roothaan equations (Section 14.3). If a large enough basis set is used, the MOs are accurate approximations to the Hartree–Fock MOs. If a minimal basis set is used, the MOs are only rough approximations to the Hartree–Fock MOs but are still referred to as SCF molecular orbitals. The SCF MO method has been widely used in polyatomic-molecule electronic-structure calculations, but is now losing ground to density functional calculations and to other methods that allow for electron correlation (Chapter 16).

How are polyatomic MOs classified? As might be expected, the MOs of a polyatomic molecule show the same kinds of possible symmetry behavior as the overall electronic wave function does (Section 15.2); for the proof, see C. C. J. Roothaan, *Rev. Mod. Phys.*,

23, 69 (1951). The MOs are therefore classified according to the symmetry species of the molecular point group. For example, the MOs of  $\text{H}_2\text{O}$  have the possible symmetry species  $a_1$ ,  $a_2$ ,  $b_1$ , and  $b_2$  of (15.3). Lowercase letters are used for MO symmetry species. To distinguish MOs of the same symmetry species, we number them in order of increasing energy. Thus the lowest three  $a_1$  MOs of water are called  $1a_1$ ,  $2a_1$ , and  $3a_1$ . This nomenclature is similar to that of the third column in Table 13.1.

Each MO holds two electrons of opposite spin. MOs having the same energy constitute a **shell**. A shell that consists of a single MO of symmetry species  $a$  or  $b$  holds two electrons. A shell that consists of two  $e$  MOs having the same energy holds four electrons; and so on. For  $\text{H}_2\text{O}$  there are no degenerate symmetry species, and each shell holds two electrons. For  $\text{C}_6\text{H}_6$  there are some doubly degenerate symmetry species [see (15.9)], so some of the benzene MOs occur in pairs having the same energy. Specification of the number of electrons in each shell specifies the molecular **electronic configuration**. Just as in atoms and diatomic molecules, a given electron configuration of a polyatomic molecule gives rise to one or more electronic terms. [For example, an  $(e_{1g})^2$  configuration of a  $\mathcal{D}_{6h}$  molecule gives the terms  $^1A_{1g}$ ,  $^1E_{2g}$ , and  $^3A_{2g}$ .] The systematic method for finding the terms of a configuration uses group theory and will not be discussed here. For tables of the terms arising from various configurations, see G. Herzberg, *Electronic Spectra of Polyatomic Molecules*, Van Nostrand, 1966, pp. 330–334, 570–573. A closed-shell configuration gives rise to a single nondegenerate term whose spin multiplicity is 1 and whose symmetry species is the totally symmetric one. Most polyatomic molecules have a closed-shell ground state, for which the MO wave function is a single Slater determinant. For states arising from open-shell configurations, the MO wave function may require a linear combination of a few Slater determinants.

### SCF MO Wave Functions for Open-Shell States

For SCF MO calculations on closed-shell states of molecules and atoms, electrons paired with each other are almost always given precisely the same spatial orbital function. A Hartree–Fock wave function in which electrons whose spins are paired occupy the same spatial orbital is called a **restricted Hartree–Fock** (RHF) wave function. (The unmodified term *Hartree–Fock wave function* is understood to mean the RHF wave function.)

Although the RHF wave function is generally used for closed-shell states, two different approaches are widely used for open-shell states. In the **restricted open-shell Hartree–Fock** (ROHF) method, electrons that are paired with each other are given the same spatial orbital function. For example, the ROHF wave function of the Li ground state is  $|1s\bar{1}s2s|$ , where the two  $1s$  electrons occupy the same spatial MO. The  $2s$  electron in this ROHF function has been given spin  $\alpha$ . Since electrons with the same spin tend to keep away from each other (Pauli repulsion; Section 10.3), the interaction between the  $2s\alpha$  and  $1s\alpha$  electrons differs from the interaction between the  $2s\alpha$  and  $1s\beta$  electrons, and it seems reasonable to give the two  $1s$  electrons slightly different spatial orbitals, which we shall call  $1s$  and  $1s'$ . This gives the **unrestricted Hartree–Fock** (UHF) wave function for the Li ground state as  $|1s\bar{1}s'2s|$  where  $1s \neq 1s'$ . In a UHF wave function, the spatial orbitals of spin- $\alpha$  electrons are allowed to differ from those of spin- $\beta$  electrons.

The UHF wave function gives a slightly lower energy than the ROHF wave function and is more useful in predicting electron-spin-resonance spectra (see Szabo and Ostlund, Section 3.8.6). The main problem with the UHF wave function is that it is not an eigenfunction of the spin operator  $\hat{S}^2$  (nor can it be made an eigenfunction of  $\hat{S}^2$  by taking a linear combination of a few UHF functions), whereas the true wave function and the ROHF wave function are eigenfunctions of  $\hat{S}^2$ . When a UHF wave function is found, one calculates  $\langle S^2 \rangle$  for the UHF function. If the deviation of  $\langle S^2 \rangle$  from  $S(S+1)\hbar^2$  is substantial, the UHF wave function should be viewed with suspicion.

## 15.4 Basis Functions

Most molecular quantum-mechanical methods, whether SCF, CI, perturbation theory (Section 16.3), coupled cluster (Section 16.4), or density functional (Section 16.5), begin the calculation with the choice of a set of basis functions  $\chi_r$ , which are used to express the MOs  $\phi_i$  as  $\phi_i = \sum_r c_{ri} \chi_r$  [Eq. (14.33)]. (Density-functional theory uses orbitals called Kohn–Sham orbitals  $\phi_i^{\text{KS}}$  that are expressed as  $\phi_i^{\text{KS}} = \sum_r c_{ri} \chi_r$ ; see Section 16.5.) The use of an adequate basis set is an essential requirement for success of the calculation.

For diatomic molecules, the basis functions are usually taken as atomic orbitals, some centered on one atom, the remainder centered on the other atom. Each AO can be represented as a linear combination of one or more Slater-type orbitals (STOs). An STO centered on atom  $a$  has the form  $N r_a^{n-1} e^{-\zeta r_a} Y_l^m(\theta_a, \phi_a)$  [Eq. (11.14)]. For nonlinear molecules, the real form of the STOs, with  $Y_l^m$  replaced by  $[(Y_l^m)^* \pm Y_l^m]/2^{1/2}$  (Section 6.6), is used. Each MO  $\phi_i$  is expressed as  $\phi_i = \sum_r c_{ri} \chi_r$ , where the  $\chi_r$ 's are the STO basis functions. We have LC-STO MOs.

For polyatomic molecules, the LC-STO method uses STOs centered on each of the atoms. The presence of more than two atoms causes difficulties in evaluating the needed integrals. For a triatomic molecule, one must deal with three-center, two-center, and one-center integrals. For a molecule with four or more atoms, one also has four-center integrals, but the number of centers in any one integral does not exceed four. Solution of the Roothaan equations requires evaluation of the electron-repulsion integrals  $(rs|tu)$  and the  $H_{rs}^{\text{core}}$  integrals [Eq. (14.41)]. If the basis functions  $\chi_r(1)$ ,  $\chi_s(1)$ ,  $\chi_t(2)$ ,  $\chi_u(2)$  are each centered on a different nucleus, then  $(rs|tu)$  is a four-center integral. The  $H_{rs}^{\text{core}}$  integrals involve either one or two centers.

For the basis set  $\chi_1, \chi_2, \dots, \chi_b$ , there are  $b$  different possibilities for each basis function in  $(rs|tu)$ , and use of the identities  $(rs|tu) = (sr|tu) = \dots$  [Eq. (14.47)] shows that there are about  $b^4/8$  different electron-repulsion integrals to be evaluated. Accurate SCF molecular calculations on small- to medium-size molecules might use from 50 to 500 basis functions, producing from 700000 to  $10^{10}$  electron-repulsion integrals. Computer evaluation of three- and four-center integrals over STO basis functions is very time consuming.

To speed up molecular integral evaluation, Boys proposed in 1950 the use of **Gaussian-type functions** (GTFs) instead of STOs for the atomic orbitals in an LCAO wave function. A **Cartesian Gaussian** centered on atom  $b$  is defined as

$$g_{ijk} = N x_b^i y_b^j z_b^k e^{-\alpha r_b^2} \quad (15.11)$$

where  $i, j$ , and  $k$  are nonnegative integers,  $\alpha$  is a positive **orbital exponent**,  $x_b, y_b, z_b$  are Cartesian coordinates with the origin at nucleus  $b$ , and  $r_b$  is the distance to nucleus  $b$ . The Cartesian-Gaussian normalization constant is

$$N = \left( \frac{2\alpha}{\pi} \right)^{3/4} \left[ \frac{(8\alpha)^{i+j+k} i! j! k!}{(2i)! (2j)! (2k)!} \right]^{1/2}$$

When  $i + j + k = 0$  (that is,  $i = 0, j = 0, k = 0$ ), the GTF is called an **s-type** Gaussian. When  $i + j + k = 1$ , we have a **p-type** Gaussian, which contains the factor  $x_b$ ,  $y_b$ , or  $z_b$ . When  $i + j + k = 2$ , we have a **d-type** Gaussian. There are six **d-type** Gaussians, with the factors  $x_b^2, y_b^2, z_b^2, x_b y_b, x_b z_b$ , and  $y_b z_b$ . If desired, five linear combinations (having the factors  $x_b y_b, x_b z_b, y_b z_b, x_b^2 - y_b^2$ , and  $3z_b^2 - r_b^2$ ) can be formed to have the same angular behavior as the five real  $3d$  AOs; a sixth combination with the factor  $x_b^2 + y_b^2 + z_b^2 = r_b^2$  is like a  $3s$  function. This sixth combination is often omitted from the basis set. Similarly, there are ten **f-type** Gaussians, and these could be combined to have the angular behavior

of the seven real  $4f$  AOs. [In general, linear combinations of Cartesian Gaussians can be formed to have the form  $Nr_b^l e^{-\alpha r_b^2} [(Y_l^m)^* \pm Y_l^m] / 2^{1/2}$ .]

Note the absence of the principal quantum number  $n$  in (15.11). Any  $s$  AO (whether  $1s$  or  $2s$  or . . . ) is represented by a linear combination of several Gaussians with different orbital exponents, each Gaussian having the form  $\exp(-\alpha r_b^2)$ ; any atomic  $p_x$  orbital is represented by a linear combination of Gaussians, each of the form  $x_b \exp(-\alpha r_b^2)$ ; and so on. The Cartesian Gaussians form a complete set.

An alternative to Cartesian Gaussians is *spherical Gaussians*, whose real form is  $Nr_b^{n-1} e^{-\alpha r_b^2} [(Y_l^m)^* \pm Y_l^m] / 2^{1/2}$ .

The behavior of the Gaussian exponential factor is shown in Fig. 4.4a, where the origin is at nucleus  $b$ . A Gaussian function does not have the desired cusp at the nucleus and hence gives a poor representation of an AO for small values of  $r_b$ . To accurately represent an AO, we must use a linear combination of several Gaussians. Therefore, an LC-GTF SCF MO calculation involves evaluation of very many more integrals than the corresponding LC-STO SCF MO calculation, since the number of two-electron integrals is proportional to the fourth power of the number of basis functions. However, Gaussian integral evaluation takes much less computer time than Slater integral evaluation. This is because the product of two Gaussian functions centered at two different points is equal to a single Gaussian centered at a third point. Thus all three- and four-center two-electron repulsion integrals are reduced to two-center integrals.

Let us discuss some of the terminology used to describe STO basis sets. A **minimal** (or **minimum**) **basis set** consists of one STO for each inner-shell and valence-shell AO of each atom (Section 14.3). For example, for  $C_2H_2$  a minimal basis set consists of  $1s$ ,  $2s$ ,  $2p_x$ ,  $2p_y$ , and  $2p_z$  AOs on each carbon and a  $1s$  STO on each hydrogen. There are five STOs on each C and one on each H, for a total of 12 basis functions. This set contains two  $s$ -type STOs and one set of  $p$ -type STOs on each carbon and one  $s$ -type STO on each hydrogen. Such a set is denoted by  $(2s1p)$  for the carbon functions and  $(1s)$  for the hydrogen functions, a notation that is further abbreviated to  $(2s1p/1s)$ . The numbers of basis functions in a minimal STO set for the first part of the periodic table are

| H, He | Li–Ne | Na–Ar | K, Ca | Sc–Kr |
|-------|-------|-------|-------|-------|
| 1     | 5     | 9     | 13    | 18    |

For each of the atoms Na to Ar, the minimal-basis AOs are  $1s$ ,  $2s$ ,  $2p_{x,y,z}$ ,  $3s$ ,  $3p_{x,y,z}$  (but not  $3d$ ).

A **double-zeta** (DZ) **basis set** is obtained by replacing each STO of a minimal basis set by two STOs that differ in their orbital exponents  $\zeta$  (zeta). (Recall that a single STO is not an accurate representation of an AO. Use of two STOs gives substantial improvement.) For example, for  $C_2H_2$  a double-zeta set consists of two  $1s$  STOs on each H, two  $1s$  STOs, two  $2s$  STOs, two  $2p_x$ , two  $2p_y$  and two  $2p_z$  STOs on each carbon, for a total of 24 basis functions; this is a  $(4s2p/2s)$  basis set. (Recall that we did a double-zeta SCF calculation on He in Section 14.3.) Since each basis function  $\chi_r$  in  $\phi_i = \sum_i c_{ri} \chi_r$  has its own independently determined variational coefficient  $c_{ri}$ , the number of variational parameters in a double-zeta-basis-set wave function is twice that in a minimal-basis-set wave function. A **triple-zeta** (TZ) **basis set** replaces each STO of a minimal basis set by three STOs that differ in their orbital exponents.

A **split-valence** (SV) **basis set** uses two (or more) STOs for each valence AO but only one STO for each inner-shell (core) AO. An SV basis set is minimal for inner-shell AOs and double zeta (or triple zeta or . . . ) for the valence AOs. Split-valence sets are called **valence double zeta** (VDZ), **valence triple-zeta** (VTZ), . . . according to the number of STOs used for each valence AO.

AOs are distorted in shape and have their centers of charge shifted upon molecule formation. To allow for this **polarization**, one adds basis-function STOs whose  $l$  quantum numbers are greater than the maximum  $l$  of the valence shell of the ground-state atom. Any such basis set is a **polarized (P)** basis set. A common example is a **double-zeta plus polarization set** (DZ + P or DZP), which typically adds to a double-zeta set a set of five 3d functions on each “first-row” and each “second-row” atom and a set of three  $2p$  functions ( $2p_x, 2p_y, 2p_z$ ) on each hydrogen atom. In the quantum-chemistry literature, Li–Ne are called the **first-row** elements, even though they are actually the second row of the periodic table. This terminology will be used in this chapter and in Chapters 16 and 17. A DZP STO basis set for  $\text{C}_2\text{H}_5\text{OSiH}_3$  is designated as  $(6s4p1d/4s2p1d/2s1p)$ , where the slashes separate the functions for atoms of different rows of the periodic table in decreasing order. For increased accuracy, higher- $l$  polarization functions can be added.

Now consider Gaussian-basis-set terminology. Instead of using the individual Gaussian functions (15.11) as basis functions, the current practice is to take each basis function as a normalized linear combination of a few Gaussians, according to

$$\chi_r = \sum_u d_{ur} g_u \quad (15.12)$$

where the  $g_u$ 's are normalized Cartesian Gaussians [Eq. (15.11)] centered on the same atom and having the same  $i, j, k$  values as one another, but different  $\alpha$ 's. The **contraction coefficients**  $d_{ur}$  are constants that are held fixed during the calculation. In (15.12),  $\chi_r$  is called a **contracted Gaussian-type function (CGTF)** and the  $g_u$ 's are called **primitive Gaussians**. By using contracted Gaussians instead of primitive Gaussians as the basis set, the number of variational coefficients to be determined is reduced, which gives large savings in computational time with little loss in accuracy if the contraction coefficients  $d_{ur}$  are well chosen.

The classifications given for STO basis sets also apply to CGTF basis sets if “STO” is replaced by “CGTF” in each definition. For example, a **minimal basis set** of contracted Gaussians consists of one contracted Gaussian function for each inner-shell AO and for each valence-shell AO. A DZ basis set has two CGTFs for each such AO, and a DZP set adds contracted Gaussians with higher  $l$  values to the DZ set, where  $l \equiv i + j + k$  in (15.11).

In molecular calculations using CGTF basis functions, the orbital exponents and contraction coefficients of the basis functions are kept fixed at the predetermined values for the basis set used. Therefore, in a CGTF minimal-basis-set calculation, there is no way for the basis functions to adjust their sizes to differing molecular environments. By using a double-zeta basis set, we allow the AO sizes to vary from one molecule to another. For example, suppose we have two  $1s$  CGTFs  $1s'$  and  $1s''$  centered on a certain H atom. Each function  $1s'$  and  $1s''$  is a linear combination of a few  $s$ -type primitive Gaussians [Eq. (15.12)]. Let the orbital exponents in the primitives of  $1s'$  be much larger than those in  $1s''$ . Then  $1s''$  is spread out over a much larger region of space than  $1s'$ . The expression for a given MO will contain the terms  $c_1 1s' + c_2 1s''$ , where the optimum  $c_1$  and  $c_2$  values are found by the SCF process. The size of the function  $c_1 1s' + c_2 1s''$  will increase as the ratio  $c_2/c_1$  increases (assuming  $c_1$  and  $c_2$  have the same signs).

By adding polarization functions, we allow the AO shapes to vary, thereby shifting charge density away from the nuclei and into the bonding regions in the molecule. For example, adding  $2p$  functions to a  $1s$  function on a hydrogen atom, we get the terms  $c_1 1s + c_2 2p_x + c_3 2p_y + c_4 2p_z$  as part of an MO. This AO will be polarized in a direction determined by the values of  $c_2, c_3$ , and  $c_4$ . For example, suppose  $c_3 = 0, c_4 = 0, c_1 > 0, c_2 > 0$ . Then, since  $2p_x$  has opposite signs on each side of the  $x$  axis, the term

$c_2 2p_x$  will cancel some of the probability density of the  $c_1 1s$  term on one side of the H atom and will augment it on the other side, thereby polarizing the  $1s$  function in the positive  $x$  direction. (Recall the discussion of hybridization in Section 13.5.) Similarly,  $p$ -type AOs can be polarized by mixing in  $d$ -type AOs.

Several methods exist to form contracted Gaussian sets. Minimal CGTF sets are often formed by fitting STOs. One starts with a minimal basis set of one STO per AO, with the STO orbital exponents fixed at values found to work well in calculations on small molecules. Each STO is then approximated as a linear combination of  $N$  Gaussian functions, where the Gaussian orbital exponents and the coefficients in the linear combination are chosen to give the best least-squares fit to the STO. Most commonly,  $N = 3$ , giving a set of CGTFs called STO-3G; this basis set is defined for the atoms H through I. Since a linear combination of three Gaussians is only an approximation to an STO, the STO-3G basis set gives results not quite as good as a minimal-basis-set STO calculation. A minimal basis set of STOs for a compound containing only first-row elements and hydrogen is denoted by  $(2s1p/1s)$ . Since each STO is replaced by a linear combination of three primitive Gaussians (which is one contracted Gaussian), the STO-3G basis set for a compound of first-row atoms and H is denoted by  $(6s3p/3s)$  contracted to  $[2s1p/1s]$ , where *parentheses indicate the primitive Gaussians and brackets indicate the contracted Gaussians*.

Suppose we want to fit a linear combination of three Cartesian GTFs to a  $1s$  STO having  $\zeta = 1$ . In atomic units, this STO is [Eqs. (11.14) and (5.101)]  $S(r; 1) = \pi^{-1/2} e^{-r}$ , where the value of the parameter  $\zeta$  is given in parentheses. From (15.11) and the following normalization-constant expression, the normalized  $s$ -type Gaussian is  $(2\alpha/\pi)^{3/4} e^{-\alpha r^2}$ . The desired normalized linear combination of three Gaussians has the form

$$G_{3N}(r; 1) = c_1(2\alpha_1/\pi)^{3/4} e^{-\alpha_1 r^2} + c_2(2\alpha_2/\pi)^{3/4} e^{-\alpha_2 r^2} + c_3(2\alpha_3/\pi)^{3/4} e^{-\alpha_3 r^2}$$

where the six parameters  $c_1, c_2, c_3, \alpha_1, \alpha_2, \alpha_3$  must be adjusted to fit the function  $S$ , and  $N$  indicates a normalized function. Using a least-squares criterion of goodness of fit and the Solver in the Excel spreadsheet, we find (Prob. 15.9)  $c_1 = 0.444615$ ,  $c_2 = 0.535336$ ,  $c_3 = 0.154340$ ,  $\alpha_1 = 0.109814$ ,  $\alpha_2 = 0.40575$ ,  $\alpha_3 = 2.22746$ .

Suppose we now want to fit a  $1s$  STO with orbital exponent  $\zeta$  with a linear combination of three GTFs. We can show (Prob. 15.10) that the correct fit is found by taking the function  $G_{3N}(r; 1)$  and replacing each orbital exponent  $\alpha_i$  by  $\zeta^2 \alpha_i$ , while leaving the coefficients  $c_i$  unchanged. The quantity  $\zeta$  is called a **scale factor**. The value  $\zeta = 1.24$  is found to work well as the STO orbital exponent of an H-atom  $1s$  basis function in SCF MO molecular calculations. Multiplication of the preceding orbital exponents by  $(1.24)^2$  gives the orbital exponents for  $\zeta = 1.24$  as 0.16885, 0.62388, and 3.42494. These orbital exponents, along with the coefficients listed above, define the STO-3G  $1s$  orbital for H. To get the STO-3G  $1s$  orbital for another atom, one uses the appropriate  $\zeta$  for that atom and multiplies the  $\zeta = 1$  orbital exponents by  $\zeta^2$ .

Another way to form contracted Gaussians is to start with atomic GTF SCF calculations. Huzinaga used a  $(9s5p)$  basis set of uncontracted Gaussians to do SCF calculations on the atoms Li–Ne. For example, for the ground state of the O atom, the optimized orbital exponents of the nine  $s$ -type basis GTFs were found to be [S. Huzinaga, *J. Chem. Phys.*, **42**, 1293 (1965)]

| $g_1$ | $g_2$ | $g_3$ | $g_4$ | $g_5$ | $g_6$ | $g_7$ | $g_8$ | $g_9$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 7817  | 1176  | 273.2 | 81.2  | 27.2  | 9.53  | 3.41  | 0.940 | 0.285 |

The expansion coefficients for the  $1s$  SCF oxygen AO were found to be

| $g_1$  | $g_2$ | $g_3$ | $g_4$ | $g_5$ | $g_6$ | $g_7$ | $g_8$   | $g_9$ |
|--------|-------|-------|-------|-------|-------|-------|---------|-------|
| 0.0012 | 0.009 | 0.043 | 0.144 | 0.356 | 0.461 | 0.140 | −0.0006 | 0.001 |

and the  $2s$  SCF AO coefficients are

| $g_1$   | $g_2$  | $g_3$  | $g_4$  | $g_5$  | $g_6$  | $g_7$  | $g_8$ | $g_9$ |
|---------|--------|--------|--------|--------|--------|--------|-------|-------|
| -0.0003 | -0.002 | -0.010 | -0.036 | -0.095 | -0.196 | -0.037 | 0.596 | 0.526 |

Suppose we want to form a valence-double-zeta (VDZ)  $[3s2p]$  set of contracted GTFs for O. We see that the  $g_1, g_2, g_3, g_4, g_5$ , and  $g_7$  coefficients are much larger for the  $1s$  AO than for the  $2s$  AO, the  $g_8$  and  $g_9$  coefficients are much larger for the  $2s$  AO than for the  $1s$  AO, and  $g_6$  makes substantial contributions to both  $1s$  and  $2s$ . We might therefore take the contracted  $1s$  basis function as

$$1s = N(0.0012g_1 + 0.009g_2 + 0.043g_3 + 0.144g_4 + 0.356g_5 + 0.461g_6 + 0.140g_7)$$

where the normalization constant  $N$  is needed because  $g_8$  and  $g_9$  have been omitted. For a VDZ set, we need two basis functions for the  $2s$  AO. These will be formed from  $g_6, g_8$ , and  $g_9$ , which are the main contributors to the  $2s$  AO. Of these three, the function  $g_9$  has the smallest orbital exponent and so falls off most slowly as  $r$  increases. ( $g_9$  is called a **diffuse** function.) The outer region of an AO changes the most upon molecule formation, and to allow for this change we can take the diffuse function  $g_9$  as one of the basis functions, giving as the  $2s$  oxygen contracted basis functions

$$2s = N'(-0.196g_6 + 0.596g_8), \quad 2s' = g_9$$

The  $2p$  and  $2p'$  CGTFs can be formed similarly.

Dunning's DZ  $[4s2p]$  and Dunning and Hay's VDZ  $[3s2p]$  contractions of Huzinaga's ( $9s5p$ ) first-row-atom AOs were often used in molecular calculations [T. H. Dunning, *J. Chem. Phys.*, **53**, 2823 (1970); T. H. Dunning and P. J. Hay in *Schaefer, Methods of Electronic Structure Theory*, pp. 1–27].

The 3-21G basis set (defined for the atoms H through Cs) and the 6-31G set (defined for H through Zn) are VDZ basis sets of CGTFs. In the 3-21G set, each inner-shell AO ( $1s$  for Li–Ne;  $1s, 2s, 2p_x, 2p_y, 2p_z$  for Na–Ar; and so on) is represented by a single CGTF that is a linear combination of three primitive Gaussians. For each valence-shell AO ( $1s$  for H;  $2s$  and the  $2p$ 's for Li–Ne; . . . ;  $4s$  and the  $4p$ 's for K, Ca, Ga–Kr;  $4s$ , the  $4p$ 's, and the five  $3d$ 's for Sc–Zn), there are two basis functions, one of which is a CGTF that is a linear combination of two Gaussian primitives and one that is a single diffuse Gaussian. The 6-31G set uses six primitives in each inner-shell CGTF and represents each valence-shell AO by one CGTF with three primitives and one Gaussian with one primitive. The orbital exponents and contraction coefficients  $d_{ur}$  in these basis sets were determined by using these basis sets to minimize the SCF energies of atoms. However, in the 3-21G basis set, the orbital exponents found for H in an atomic calculation are increased using a scale factor, and in the 6-31G set, the valence orbital exponents of H and of Li through O are scaled to be more appropriate for molecular calculations.

The 6-31G\* basis set (defined for the atoms H through Zn) is a valence double-zeta polarized basis set that adds to the 6-31G set six  $d$ -type Cartesian–Gaussian polarization functions on each of the atoms Li through Ca and ten  $f$ -type Cartesian–Gaussian polarization functions on each of the atoms Sc through Zn. The 6-31G\*\* basis set adds to the 6-31G\* set a set of three  $p$ -type Gaussian polarization functions on each hydrogen and helium atom. The orbital exponents of the polarization functions in these two basis sets were determined as the average of the optimum values found in calculations on small molecules. Each polarization function in 6-31G\* or 6-31G\*\* consists of a single primitive Gaussian function. The 6-31G\* and 6-31G\*\* sets are sometimes denoted as 6-31G( $d$ ) and 6-31G( $d, p$ ), respectively. In the 6-31G\* basis set, a phosphorus atom has 19 basis functions centered on it ( $1s, 2s, 2p_x, 2p_y, 2p_z, 3s, 3s', 3p_x, 3p_y, 3p_z, 3p'_x, 3p'_y, 3p'_z$ , and six  $d$ 's) and is  $[4s3p1d]$  for P.

For second-row atoms,  $d$  orbitals contribute significantly to the bonding. To allow for this, the 3-21G(\*) basis set (defined for H through Ar) is constructed by the addition to the 3-21G set of a set of six  $d$ -type Gaussian functions on each second-row atom (Na through Ar). For H–Ne, the 3-21G(\*) set (which is often called 3-21G\*) is the same as the 3-21G set.

Anions, compounds with lone pairs, and hydrogen-bonded dimers have significant electron density at large distances from the nuclei. To improve the accuracy for such compounds, the 3-21+G and 6-31+G\* basis sets are formed from the 3-21G and 6-31G\* sets by the addition of four highly diffuse functions ( $s, p_x, p_y, p_z$ ) on each nonhydrogen atom. A **highly diffuse function** is one with a very small orbital exponent (typically, 0.01 to 0.1). The 3-21++G and 6-31++G\* sets also include a highly diffuse  $s$  function on each hydrogen atom.

Basis sets larger than the relatively small 6-31G\* and 6-31G\*\* basis sets are often used in calculations that include electron correlation. An example is the 6-311G\*\* set defined for H–Ca, Ga–Kr and I, which is single zeta for the core AOs, triple zeta for the valence AOs, and contains five  $d$ -type Gaussian polarization functions on each first- and second-row atom and three  $p$ -type polarization functions on each hydrogen atom. The 6-311G\*\* set cannot really be considered to be a large basis set. Large basis sets typically use more than one set of polarization functions on each atom. An example of a large basis set for correlation calculations is the 6-311++G(3df, 3pd) set, defined for H–Ar. The pluses indicate diffuse functions on all atoms; the letters in parentheses indicate that three sets of five  $d$ -type Gaussian polarization functions (each set having a different orbital exponent) are added to each nonhydrogen (giving fifteen  $d$  functions on nonhydrogens), one set of seven  $f$ -type Gaussians is added to each nonhydrogen, and three sets of three  $p$ -type Gaussians and one set of five  $d$ -type Gaussians are added to each hydrogen. This is a VTZ3PD set, meaning valence triple zeta, triple polar, and diffuse functions on all atoms.

The basis sets STO-3G, 3-21G, 3-21G(\*), 6-31G\*, 6-31G\*\*, etc., were developed by Pople and co-workers (see *Hehre et al.*, Section 4.3).

Dunning and co-workers have developed the CGTF basis sets cc-pVDZ, cc-pVTZ, cc-pVQZ, cc-pV5Z, and cc-pV6Z (collectively denoted as cc-pVnZ, where  $n$  goes from 2 to 6), designed for use in calculation methods (such as CI) that include electron correlation. Here, cc-pVDZ stands for *correlation-consistent, polarized valence double-zeta*. Unlike the Pople-type functions, where the number of  $d$  functions used varies from set to set, the cc family of functions always uses five  $d$  functions, seven  $f$  functions, etc. These sets are defined for the elements H–Ar, and Ca–Kr for  $n = 2, 3, 4$ , and 5; and for H, He, B–Ne, and Al–Ar for  $n = 6$ . For first-row atoms, the contracted Gaussians present in the cc basis sets are

| cc-pVDZ  | cc-pVTZ    | cc-pVQZ      | cc-pV5Z        | cc-pV6Z          |
|----------|------------|--------------|----------------|------------------|
| [3s2p1d] | [4s3p2d1f] | [5s4p3d2f1g] | [6s5p4d3f2g1h] | [7s6p5d4f3g2h1i] |
| 14       | 30         | 55           | 91             | 140              |

where the last row gives the number of basis functions for a first-row atom [for example,  $3(1) + 2(3) + 1(5) = 14$  for cc-pVDZ]. Note that as we go from one basis set to the next, the number of sets of basis functions of each angular-momentum  $l$  value is increased by one and one set of functions with the next higher  $l$  value is added. The idea is that in going from one set to the next larger set, one adds basis functions that make similar contributions to the correlation energy (hence the name “correlation consistent”). For example, the first set of  $f$  functions has been found to lower the energy by a similar amount as the

second set of added  $d$  functions, so in going from cc-pVDZ to cc-pVTZ, one adds a second set of  $d$  functions and the first set of  $f$  functions (and an additional  $s$  and a  $p$  set).

For an H atom, the cc-pVDZ set is  $[2s1p]$  and the same pattern is followed as for first-row atoms. For example, in going to cc-pVTZ for H, one adds an  $s$  set, a  $p$  set, and a  $d$  set to get  $[3s2p1d]$ .

The cc-pVDZ set is roughly comparable to the 6-31G\*\* set, since both are  $[2s1p]$  for H and  $[3s2p1d]$  for first-row atoms.

The addition of diffuse primitive nonpolarization and polarization functions to the cc-pVnZ basis sets gives the augmented sets aug-cc-pVDZ, aug-cc-pVTZ, etc., especially suitable for calculations on anions and hydrogen-bonded species. To form the set aug-cc-pVnZ from cc-pVnZ, the number of sets of basis functions of each angular-momentum  $l$  value is increased by one by the addition of diffuse primitives. Thus, aug-cc-pVTZ is  $[5s4p3d2f]$  for a first-row atom. For calculation of the electric polarizability of molecules, the convergence rate with increase in basis-set size is greatly increased by using doubly augmented (d-aug-cc-pVnZ) sets. For a first-row atom, d-aug-cc-pVTZ is  $[6s5p4d3f]$ .

The addition of certain primitive Gaussians to the cc-pVnZ sets gives the cc-pCVDZ, cc-pCVTZ, . . . sets (where CV stands for core/valence), which are designed for calculations that include correlation effects involving the core electrons. Diffuse functions can be added to these CV sets to give the aug-cc-pCVnZ sets.

The correlation-consistent basis sets have the advantage that as the basis-set size is expanded, the calculated values of a molecular property usually converge smoothly to a limiting value, making it easy to find this value by extrapolation. Because of this feature, these basis sets are the most widely used sets in high-level ab initio calculations. The disadvantage of these basis sets is that they are rather large.

A review of what degree of augmentation with diffuse basis functions is needed for calculations of various molecular properties by various methods is given in E. Papajak et al., *J. Chem. Theory Comput.*, **7**, 3027 (2011). These workers found that augmentation with diffuse functions is generally not needed for hydrogen atoms. Deletion of H-atom diffuse functions from a basis set gives a basis set denoted by the prefix jul- by these workers, where jul is used because July precedes August in the calendar. Thus the jul-cc-pV5Z basis set is formed from the aug-cc-pV5Z set by deleting the diffuse functions on the H atoms. Deletion of successive diffuse subshells on heavier atoms gives the jun-, may-, and apr- calendar basis sets. For example, the aug-cc-pVTZ basis set has  $s$ ,  $p$ ,  $d$ , and  $f$  diffuse functions on each carbon atom, and the jun-cc-pVTZ set deletes the  $f$  diffuse functions on C and the diffuse functions on H.

Commonly used Gaussian basis sets and literature references for them are available at the **Basis Set Exchange** website ([bse.pnl.gov/bse/portal](http://bse.pnl.gov/bse/portal)) at the Environmental Molecular Sciences Laboratory of the Pacific Northwest Laboratory.

Because of the time savings given by Gaussians in multicenter-integral evaluation, nearly all current molecular ab initio calculations use contracted-Gaussian basis sets. Most semiempirical methods (which neglect large classes of integrals) use STOs. Density-functional calculations may use Gaussians, STOs, or other basis functions.

#### EXAMPLE

Use the Basis Set Exchange on the Internet to find the 3-21G basis functions for the oxygen atom.

On the basis-set form, choose the 3-21G basis set and click O on the periodic table. From the drop-down menu at the upper left choose All Electron. The website allows one to choose the basis-set output to correspond to the format used by one of several programs. If we choose *Gaussian94* and click on Get Basis Set, we get [Table 15.1](#). The

**TABLE 15.1 The 3-21G Basis Set for the Oxygen Atom**

|    |              |             |            |
|----|--------------|-------------|------------|
| O  | 0            |             |            |
| S  | 3            | 1.00        |            |
|    | 322.03700000 | 0.05923940  |            |
|    | 48.43080000  | 0.35150000  |            |
|    | 10.42060000  | 0.70765800  |            |
| SP | 2            | 1.00        |            |
|    | 7.40294000   | −0.40445300 | 0.24458600 |
|    | 1.57620000   | 1.22156000  | 0.85395500 |
| SP | 1            | 1.00        |            |
|    | 0.37368400   | 1.00000000  | 1.00000000 |

O in the first row indicates the oxygen atom. The zero after the O is required by the *Gaussian 94* program and can be ignored. On the second line, the S and 3 indicate that an *s*-type CGTF consisting of 3 primitive Gaussians follows; the 1.00 is a scale factor and since its value is 1, it can be ignored. The first and second columns in the next three rows give the orbital exponents and contraction coefficients, respectively. The large values of these orbital exponents show that this is the inner-shell (core) 1*s* AO. Thus the 1*s* CGTO is

$$1s = 0.0592394g_s(322.037) + 0.3515g_s(48.4308) + 0.707658g_s(10.4206)$$

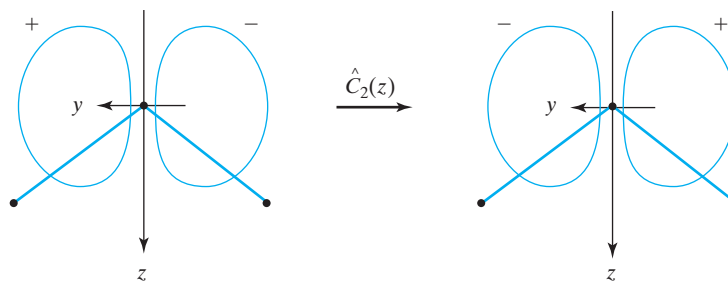
where  $g_s(322.037)$  denotes a normalized primitive *s*-type GTF with orbital exponent 322.037 [Eq. (15.11)]. The SP on the sixth and ninth lines indicates that orbital exponents and contraction coefficients for *s*-type and *p*-type CGTFs follow. These are for the valence 2*s* and 2*p* AOs. The 3-21G set uses the same orbital exponents for the 2*s* and 2*p* AOs, so as to speed up calculations. The first column of numbers gives the orbital exponents and the second and third columns give the contraction coefficients. Thus the valence CGTFs are  $2s' = -0.404453g_s(7.40294) + 1.22156g_s(1.5762)$ ,  $2p'_x = 0.244586g_{p_x}(7.40294) + 0.853955g_{p_x}(1.5762)$ ,  $\dots$ ,  $2s'' = g_s(0.373684)$ ,  $2p''_x = g_{p_x}(0.373684)$ ,  $\dots$ , where the dots indicate  $2p_y$  and  $2p_z$  CGTFs. [Sometimes people omit the normalization constant from each primitive Gaussian (15.11) and include it within the contraction coefficient for that Gaussian. The fact that the contraction coefficient for each outer 2*s* and 2*p* CGTF is 1.00 assures us that the primitive Gaussians are normalized.]

To describe a quantum-mechanical calculation, one specifies the method and the basis set. The letters HF (for Hartree–Fock) denote any ab initio SCF MO calculation, whether or not the basis set is large enough to come close to the Hartree–Fock limit. Thus the notation HF/3-21G denotes an ab initio SCF MO calculation that uses the 3-21G basis set.

## 15.5 The SCF MO Treatment of H<sub>2</sub>O

For a minimal-basis-set MO treatment of H<sub>2</sub>O (Fig. 15.1 in Section 15.2), we start with the O1*s*, O2*s*, O2*p<sub>x</sub>*, O2*p<sub>y</sub>*, and O2*p<sub>z</sub>* inner-shell and valence oxygen AOs, and the H<sub>1</sub>1*s* and H<sub>2</sub>1*s* valence AOs of the hydrogen atoms. Linear combinations of these seven basis AOs give LCAO approximations to the seven lowest MOs of water. As stated in Section 15.3, the MOs can be chosen so that upon application of the molecular symmetry operators each

**FIGURE 15.2** The effect of  $\hat{C}_2(z)$  on the  $2p_y$  oxygen AO in  $\text{H}_2\text{O}$ .



MO transforms according to one of the symmetry species of the molecular point group. To aid in choosing the right linear combinations of AOs, we examine the symmetry behavior of the AOs.

The  $1s$  and  $2s$  oxygen AOs are spherically symmetric, so rotation about the  $C_2(z)$  axis and reflection in the  $xz$  or  $yz$  plane has no effect on them. They thus belong to the totally symmetric symmetry species  $a_1$  of (15.3). The effect of a  $\hat{C}_2(z)$  rotation on the oxygen  $2p_y$  AO is shown in Fig. 15.2. This rotation sends  $\text{O}2p_y$  into its negative (see also Fig. 12.10). Reflection in the  $xz$  plane sends  $\text{O}2p_y$  into its negative, and reflection in the molecular  $yz$  plane sends  $\text{O}2p_y$  into itself. The symmetry eigenvalues of  $\text{O}2p_y$  are  $1, -1, -1$ , and  $1$ . The symmetry species of this AO is  $b_2$ . Similarly, the symmetry species of  $\text{O}2p_x$  and  $\text{O}2p_z$  are found to be  $b_1$  and  $a_1$ , respectively.

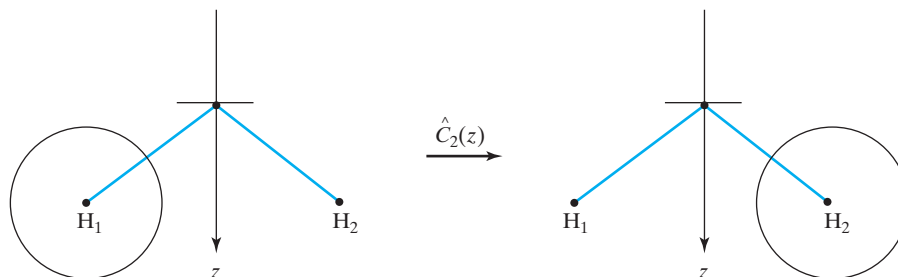
Now for the hydrogen  $1s$  AOs. Reflection in the  $yz$  plane leaves each of them unchanged. However, rotation by  $180^\circ$  about the  $z$  axis sends  $\text{H}_1 1s$  into  $\text{H}_2 1s$  and vice versa (Fig. 15.3). The  $\text{H}_1 1s$  and  $\text{H}_2 1s$  functions are *not* eigenfunctions of  $\hat{O}_{C_2(z)}$  and so do not transform according to any of the symmetry species of  $\text{H}_2\text{O}$ .

As a preliminary step in finding the MOs of a molecule, it is helpful (but not essential) to construct linear combinations of the original basis AOs such that each linear combination does transform according to one of the molecular symmetry species. Such linear combinations are called **symmetry orbitals** or **symmetry-adapted basis functions**. The symmetry orbitals are used as the basis functions  $\chi_s$  in the expansions  $\phi_i = \sum_s c_{si} \chi_s$  [Eq. (14.33)] of the MOs  $\phi_i$ . *The use of basis functions that transform according to the molecular symmetry species simplifies the calculation by putting the secular determinant in block-diagonal form.* This will be illustrated below.

Each oxygen AO transforms according to one of the symmetry species of  $\text{H}_2\text{O}$  and can serve as a symmetry orbital. However, neither of the two hydrogen  $1s$  AOs belongs to a symmetry species of  $\text{H}_2\text{O}$ , and we must construct two symmetry orbitals from these AOs. Consider the linear combinations:

$$\text{H}_1 1s + \text{H}_2 1s \quad \text{and} \quad \text{H}_1 1s - \text{H}_2 1s \quad (15.13)$$

**FIGURE 15.3** The effect of a  $\hat{C}_2(z)$  rotation on the  $\text{H}_1 1s$  AO in  $\text{H}_2\text{O}$ .



We have

$$\hat{O}_{C_2(z)}(H_1 1s + H_2 1s) = H_2 1s + H_1 1s$$

$$\hat{O}_{C_2(z)}(H_1 1s - H_2 1s) = H_2 1s - H_1 1s$$

Thus the first and second functions in (15.13) are eigenfunctions of  $\hat{O}_{C_2(z)}$  with eigenvalues  $+1$  and  $-1$ , respectively. Examination of the effects of the other three symmetry operators shows the functions (15.13) to belong to the symmetry species  $a_1$  and  $b_2$ , respectively. We shall not bother to normalize the symmetry orbitals (15.13). The seven basis symmetry functions and their symmetry species are then

|                   |          |          |          |                   |          |          |         |
|-------------------|----------|----------|----------|-------------------|----------|----------|---------|
| $\chi_1$          | $\chi_2$ | $\chi_3$ | $\chi_4$ | $\chi_5$          | $\chi_6$ | $\chi_7$ | (15.14) |
| $H_1 1s + H_2 1s$ | $O 1s$   | $O 2s$   | $O 2p_z$ | $H_1 1s - H_2 1s$ | $O 2p_y$ | $O 2p_x$ |         |
| $a_1$             | $a_1$    | $a_1$    | $a_1$    | $b_2$             | $b_2$    | $b_1$    |         |

Now consider the SCF secular determinant  $\det(F_{rs} - \epsilon_i S_{rs})$  [Eq. (14.36)]. We assert that

$$F_{rs} \equiv \langle \chi_r | \hat{F} | \chi_s \rangle = 0 \quad (15.15)$$

whenever  $\chi_r$  and  $\chi_s$  belong to different symmetry species. This result follows from the theorem [Eq. (7.50)] that  $\langle g_j | \hat{B} | g_k \rangle = 0$  if  $g_j$  and  $g_k$  are eigenfunctions of a Hermitian operator  $\hat{A}$  with different eigenvalues, where  $\hat{A}$  commutes with  $\hat{B}$ . The symmetry orbitals of H<sub>2</sub>O are eigenfunctions of the symmetry operators, each of which commutes with the electronic Hamiltonian and with the Fock operator  $\hat{F}$ . Symmetry orbitals  $\chi_r$  and  $\chi_s$  that belong to different symmetry species differ in at least one symmetry eigenvalue. Hence (15.15) follows. Moreover, since two eigenfunctions of a Hermitian operator that correspond to different eigenvalues are orthogonal, we have

$$S_{rs} \equiv \langle \chi_r | \chi_s \rangle = 0 \quad (15.16)$$

whenever  $\chi_r$  and  $\chi_s$  belong to different symmetry species. From (15.15) and (15.16), it follows that the use of symmetry orbitals puts the secular determinant of H<sub>2</sub>O in block-diagonal form, each block corresponding to a different symmetry species. The blocks are  $4 \times 4$ ,  $2 \times 2$ , and  $1 \times 1$ . [For molecules with degenerate symmetry species ( $E$ ,  $T$ , and so on), the symmetry orbitals of the degenerate species are not necessarily eigenfunctions of the symmetry operators. Nevertheless, the symmetry orbitals still put the secular determinant in block-diagonal form, as can be shown using group theory.]

The set of Roothaan simultaneous equations (14.34) then breaks up into one set of four simultaneous equations, one set of two simultaneous equations, and one set of one “simultaneous” equation (Section 9.6). The first set contains matrix elements involving only the four  $a_1$  symmetry orbitals. Therefore, four of the lowest seven H<sub>2</sub>O MOs are linear combinations of the four  $a_1$  symmetry orbitals. These four MOs must have  $a_1$  symmetry. Similarly, we have two MOs of  $b_2$  symmetry and one MO of  $b_1$  symmetry. The symmetry orbitals are *not* (in general) the MOs, but *each MO must be a linear combination of those symmetry orbitals having the same symmetry species as the MO*. The forms of the lowest MOs of H<sub>2</sub>O are then

$$\begin{aligned}
 \phi_1 &= c_{11}\chi_1 + c_{21}\chi_2 + c_{31}\chi_3 + c_{41}\chi_4 & \phi_5 &= c_{55}\chi_5 + c_{65}\chi_6 \\
 \phi_2 &= c_{12}\chi_1 + c_{22}\chi_2 + c_{32}\chi_3 + c_{42}\chi_4 & \phi_6 &= c_{56}\chi_5 + c_{66}\chi_6 \\
 \phi_3 &= c_{13}\chi_1 + c_{23}\chi_2 + c_{33}\chi_3 + c_{43}\chi_4 & \phi_7 &= \chi_7 \\
 \phi_4 &= c_{14}\chi_1 + c_{24}\chi_2 + c_{34}\chi_3 + c_{44}\chi_4
 \end{aligned} \quad (15.17)$$

The next step in the SCF MO calculation is to choose explicit forms for the seven AOs. The orbital energies and the coefficients of the symmetry orbitals are then found using Roothaan's equations.

Pitzer and Merrifield did an H<sub>2</sub>O minimal-basis-set calculation, representing each AO by a single STO [R. M. Pitzer and D. P. Merrifield, *J. Chem. Phys.*, **52**, 4782 (1970); S. Aung, R. M. Pitzer, and S. I. Chan, *J. Chem. Phys.*, **49**, 2071 (1968)]. They optimized the orbital exponents, finding 1.27 for H1s, 7.66 for O1s, 2.25 for O2s, and 2.21 for O2p. (To optimize the exponents, one must repeat the entire SCF iterative calculation for several different sets of orbital exponents to locate the set that gives the minimum energy. Since orbital-exponent optimization is time consuming, it is only feasible for small molecules.) The calculated orbital energies in hartrees are:  $1a_1$ , -20.56;  $2a_1$ , -1.28;  $1b_2$ , -0.62;  $3a_1$ , -0.47;  $1b_1$ , -0.40. The ground-state electron configuration of this ten-electron molecule is

$$(1a_1)^2(2a_1)^2(1b_2)^2(3a_1)^2(1b_1)^2 \quad (15.18)$$

The ground state has a closed-shell configuration and is a  $^1A_1$  state.

The five lowest SCF MOs found by Pitzer and Merrifield at the experimental geometry are

$$\begin{aligned} 1a_1 &= 1.000(\text{O}1s) + 0.015(\text{O}2s_{\perp}) + 0.003(\text{O}2p_z) - 0.004(\text{H}_11s + \text{H}_21s) \\ 2a_1 &= -0.027(\text{O}1s) + 0.820(\text{O}2s_{\perp}) + 0.132(\text{O}2p_z) + 0.152(\text{H}_11s + \text{H}_21s) \\ 1b_2 &= 0.624(\text{O}2p_y) + 0.424(\text{H}_11s - \text{H}_21s) \\ 3a_1 &= -0.026(\text{O}1s) - 0.502(\text{O}2s_{\perp}) + 0.787(\text{O}2p_z) + 0.264(\text{H}_11s + \text{H}_21s) \\ 1b_1 &= \text{O}2p_x \end{aligned} \quad (15.19)$$

The  $\text{O}2s_{\perp}$  orbital in (15.19) is an orthogonalized orbital [Eq. (13.125)]:

$$\text{O}2s_{\perp} = 1.028[\text{O}2s - 0.2313(\text{O}1s)] \quad (15.20)$$

where  $\text{O}2s$  is the ordinary  $2s$  STO:

$$\text{O}2s = 2.25^{5/2}\pi^{-1/2}3^{-1/2}r_O \exp(-2.25r_O)$$

where  $r_O$  is the distance to the oxygen nucleus. The MO approximation to the ground-state wave function of H<sub>2</sub>O is the  $10 \times 10$  Slater determinant

$$|1a_1\overline{1a_1}2a_1\overline{2a_1}1b_2\overline{1b_2}3a_1\overline{3a_1}1b_1\overline{1b_1}| \quad (15.21)$$

Consider the nature of the MOs. The lowest MO,  $1a_1$ , is essentially a pure nonbonding  $1s$  oxygen AO, which is hardly surprising.

The oxygen part of the  $2a_1$  MO is mostly  $\text{O}2s$  with some  $\text{O}2p_z$  mixed in. This mixing in (hybridization) of  $\text{O}2p_z$  adds to the value of the  $\text{O}2s$  AO along the positive  $z$  axis (which lies between the hydrogens; Fig. 15.1) and subtracts from  $\text{O}2s$  along the negative  $z$  axis. The combination of the hybridized  $\text{O}2s$  and  $\text{O}2p_z$  orbitals with the  $\text{H}_11s$  and  $\text{H}_21s$  orbitals in the  $2a_1$  MO then gives electron probability-density buildup in the region enclosed by the three nuclei. Therefore, the  $2a_1$  MO contributes to the bonding in water.

Consider the  $1b_2$  MO. The  $2p_y$  oxygen AO has its positive lobe on the  $\text{H}_1$  side of the molecule, and so the positive lobe of  $\text{O}2p_y$  adds to  $\text{H}_11s$  in the  $1b_2$  MO, giving electron charge buildup between the  $\text{H}_1$  and O nuclei. Similarly, the negative lobe of  $\text{O}2p_y$  adds to  $-\text{H}_21s$ , giving charge buildup between O and  $\text{H}_2$  in this MO. Hence  $1b_2$  is a bonding MO.

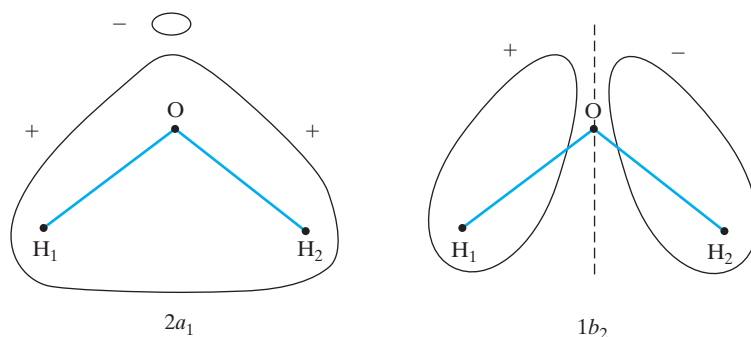


**FIGURE 15.4** Sketches of the two main bonding MOs of H<sub>2</sub>O.

The hybridization of the  $2s$  and  $2p_z$  oxygen AOs in  $3a_1$  builds up electron probability density in the region around the negative  $z$  axis, away from the hydrogens, giving this MO substantial lone-pair character. On the positive side of the  $z$  axis, the oxygen  $2s$  and  $2p_z$  AOs tend to cancel each other, and we get little bonding overlap between the oxygen hybrid and the hydrogen AOs. Alternatively, we can look separately at the overlap between  $O2s_{\perp}$  and the hydrogens (which is negative or antibonding) and the overlap between  $O2p_z$  and the hydrogens (which is positive or bonding). Because of the approximate cancellation of these overlaps (see Section 15.6), the  $3a_1$  MO has little bonding character and is best described as mainly a lone-pair MO.

The  $1b_1$  MO is a nonbonding lone-pair  $2p_x$  oxygen AO.

**Figure 15.4** shows the shapes of the bonding MOs  $2a_1$  and  $1b_2$ . Note that their symmetry eigenvalues are as given by (15.3). [For accurately plotted H<sub>2</sub>O MO contours, see T. H. Dunning, R. M. Pitzer, and S. Aung, *J. Chem. Phys.*, **57**, 5044 (1972).]

Each of the SCF bonding MOs in (15.19) and Fig. 15.4 is delocalized over the entire molecule and does not resemble a chemical bond. The relation of these MOs to the picture presented in many chemistry books, where one bonding MO points along the O—H<sub>1</sub> bond and the other points along the O—H<sub>2</sub> bond, is discussed in Section 15.8.

The unoccupied  $4a_1$  and  $2b_2$  MOs of water calculated by Pitzer and Merrifield (for nonoptimized, Slater-rule exponents) are

$$\begin{aligned} 4a_1 &= 0.08(O1s) + 0.84(O2s_{\perp}) + 0.70(O2p_z) - 0.75(H_11s + H_21s) \\ 2b_2 &= 0.99(O2p_y) - 0.89(H_11s - H_21s) \end{aligned} \quad (15.22)$$

For these two MOs, the opposite signs of the oxygen and hydrogen AO coefficients give charge depletion between the nuclei. These MOs are antibonding.

The unoccupied orbitals (15.22) are called **virtual orbitals**. Because they were calculated for the electron configuration (15.18), they are not accurate representations of the SCF orbitals actually occupied in H<sub>2</sub>O excited electronic states. When the electron configuration changes, the interorbital electronic interactions change, thereby changing the forms of all the SCF orbitals. We can, however, use the virtual orbitals and their calculated energies as rough approximations to the higher SCF orbitals and energies.

The formation of the H<sub>2</sub>O MOs from the separated-atoms AOs is illustrated schematically in **Fig. 15.5**. The dashed lines indicate which AOs contribute significantly to each MO. Note that only AOs of roughly the same energy combine to a significant degree in a given MO. [This fact is explained by the  $1/(E_n^{(0)} - E_m^{(0)})$  term in Eq. (9.28).] We can thus get a qualitative idea of the MOs without doing any calculations by using the rules that *only symmetry orbitals of the same symmetry species combine* and that *only AOs of comparable energy* (Fig. 11.2) *contribute significantly to a given MO*.

**FIGURE 15.5** Formation of the H<sub>2</sub>O MOs from the minimal-basis AOs. The five lowest MOs are filled in the ground state. (Note the break in the scale.)

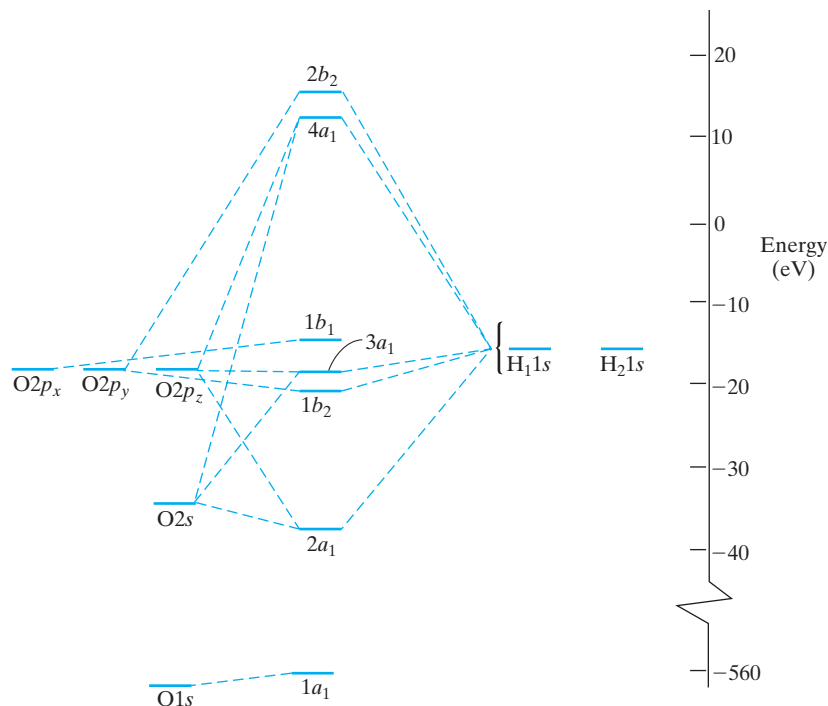


**Table 15.2** lists some of the many ab initio SCF MO calculations on the H<sub>2</sub>O ground electronic state. The size and quality of the basis set used determine how closely the Hartree–Fock limit is approached. The experimental equilibrium electronic energy of H<sub>2</sub>O is  $-76.480$  hartrees. (This is the energy needed to go from the molecule at its equilibrium geometry to nuclei and electrons at infinite separation from one another.) All the calculations listed are nonrelativistic. The only significant relativistic correction will occur in the  $1s$  inner-shell electrons of oxygen (see Section 11.7). Since this shell remains essentially unchanged on molecule formation, one can use the relativistic energy correction calculated for the O atom as the relativistic correction for H<sub>2</sub>O. In addition, there is a small correction due to motion of the center of mass of the nuclei relative to the center of mass of the molecule. When these two corrections are subtracted from the experimental energy of  $-76.480$  hartrees, we obtain the nonrelativistic fixed-nuclei H<sub>2</sub>O energy as  $-76.438$  hartrees.

Note that, except for the first three calculations in Table 15.2, the SCF MO calculations give a bond length that is somewhat shorter than the true length. This is a general characteristic of the HF method, as is the tendency of HF calculations to predict a dipole moment that is somewhat too large.

As mentioned in Section 15.4, the cc basis sets can be used to extrapolate to the complete-basis-set (CBS) limit for molecular properties. Energies in hartrees for SCF MO H<sub>2</sub>O optimized-geometry calculations with the cc-pVnZ basis sets for  $n = 2, 3, 4, 5, 6$  are  $-76.02705$ ,  $-76.05777$ ,  $-76.06552$ ,  $-76.06778$ ,  $-76.06810$ , respectively. The following empirical formula has been found to work rather well for extrapolating cc-pVnZ SCF MO energies to the CBS limit:

$$E_{\text{SCF}}(n) = E_{\text{SCF}}(\infty) + Ae^{-Bn} \quad (15.23)$$

where  $E_{\text{SCF}}(n)$  is the SCF energy found with the cc-pVnZ set,  $E_{\text{SCF}}(\infty)$  is the CBS limit predicted by cc-pVnZ in the limit  $n \rightarrow \infty$ , and  $A$  and  $B$  are positive parameters whose values

**TABLE 15.2 H<sub>2</sub>O Hartree–Fock Calculations<sup>a</sup>**

| Reference <sup>b</sup>                                 | Basis Set <sup>c</sup>            | Energy/ $E_h$        | $\mu/D$ | $\theta$            | $R_{OH}/\text{\AA}$ |
|--------------------------------------------------------|-----------------------------------|----------------------|---------|---------------------|---------------------|
| CCCBDB                                                 | STO-3G, 7                         | −74.966              | 1.71    | 100.0°              | 0.989               |
| CCCBDB                                                 | 3-21G, 13                         | −75.586              | 2.39    | 107.7°              | 0.967               |
| Pitzer, Merrifield                                     | Minimal STO, 7                    | −75.705              | 1.92    | 100.3°              | 0.990               |
| CCCBDB                                                 | 6-31G*, 19                        | −76.011              | 2.20    | 105.5°              | 0.947               |
| CCCBDB                                                 | 6-31G**, 25                       | −76.024              | 2.15    | 106.0°              | 0.943               |
| CCCBDB                                                 | cc-pVDZ, 24                       | −76.027              | 2.04    | 104.6°              | 0.946               |
| K. S. Kim et al.                                       | 6-311++G(2d,2p), 47               | −76.057              | 2.02    | 106.3°              | 0.940               |
| CCCBDB                                                 | cc-pVTZ, 58                       | −76.058              | 1.99    | 106.0°              | 0.941               |
| CCCBDB                                                 | aug-cc-pVTZ, 92                   | −76.061              | 1.93    | 106.3°              | 0.941               |
| Dunning et al.                                         | [6s5p2d/3s1p], 43                 | −76.062              | 2.08    | 106.6°              | 0.941               |
| Rosenberg et al.                                       | (5s4p2d/3s1p) <sub>STO</sub> , 39 | −76.064              | 2.00    | 106.1°              | 0.940               |
| CCCBDB                                                 | cc-pVQZ, 115                      | −76.0655             | 1.97    | 106.2°              | 0.940               |
| CCCBDB                                                 | aug-cc-pVQZ, 172                  | −76.0667             | 1.94    | 106.3°              | 0.940               |
| Amos                                                   | [8s6p4d2f/6s3p1d], 112            | −76.0675             |         | 106.3°              | 0.940               |
| K. S. Kim et al.                                       | (13s8p4d2f/8s4p2d), 131           | −76.0676             | 1.94    | 106.3°              | 0.940               |
| Bakken et al.                                          | cc-pV5Z, 201                      | −76.0678             | 1.96    | 106.33°             | 0.9396              |
| Bakken et al.                                          | aug-cc-pV5Z, 287                  | −76.0680             | 1.94    | 106.34°             | 0.9396              |
| Cortez et al.                                          | cc-pV6Z, 322                      | −76.06810            | 1.95    | 106.34°             | 0.9396              |
|                                                        | aug-cc-pV6Z, 443                  | −76.06815            | 1.94    | 106.34°             | 0.9396              |
| Pahl, Handy                                            | special <sup>c</sup>              | −76.06817            |         | 106.34°             | 0.9396              |
| <i>Estimated Hartree–Fock energy<sup>d</sup></i>       |                                   | −76.068 <sub>3</sub> |         |                     |                     |
| <i>Nonrelativistic fixed–nuclei energy<sup>e</sup></i> |                                   | −76.438              |         |                     |                     |
| <i>Experimental values</i>                             |                                   | −76.480 <sup>e</sup> | 1.85    | 104.5 <sup>of</sup> | 0.9578 <sup>f</sup> |

<sup>a</sup>Energy/ $E_h$  is the total electronic energy including nuclear repulsion in hartrees at the calculated equilibrium geometry;  $\mu$ ,  $\theta$ , and  $R_{OH}$  are the calculated electric dipole moment, equilibrium bond angle, and equilibrium bond length.

<sup>b</sup>CCCBDB is the Computational Chemistry Comparison and Benchmark DataBase (Section 15.1); R. M. Pitzer and D. P. Merrifield, *J. Chem. Phys.*, **52**, 4782 (1970); K. S. Kim et al., *J. Chem. Phys.*, **97**, 6649 (1992); T. H. Dunning, R. M. Pitzer, and S. Aung, *J. Chem. Phys.*, **57**, 5044 (1972); B. J. Rosenberg and I. Shavitt, *J. Chem. Phys.*, **63**, 2162 (1975) and B. J. Rosenberg, W. C. Ermler, and I. Shavitt, *J. Chem. Phys.*, **65**, 4072 (1976); R. D. Amos, *J. Chem. Soc. Faraday Trans. 2*, **83**, 1595 (1987); K. S. Kim et al., op. cit; V. Bakken et al., *Mol. Phys.*, **96**, 653 (1999); M. H. Cortez et al., *J. Chem. Theory Comput.*, **3**, 1267 (2007); F. A. Pahl and N. C. Handy, *Mol. Phys.*, **100**, 3199 (2002).

<sup>c</sup>The number of basis functions is given. All the basis sets are GTFs except the Pitzer–Merrifield, Rosenberg et al., and Pahl–Handy sets. Brackets denote CGTF sets and parentheses denote an uncontracted set. A slash separates the oxygen and hydrogen basis functions. The Pahl–Handy set consists of plane waves and radial polynomials.

<sup>d</sup>This estimate is for the calculated Hartree–Fock equilibrium geometry. See D. Feller, C. M. Boyle, and E. R. Davidson, *J. Chem. Phys.*, **86**, 3424 (1987).

<sup>e</sup>A. Lüchow, J. B. Anderson, and D. Feller, *J. Chem. Phys.*, **106**, 7706 (1997).

<sup>f</sup>P. Jensen et al., *J. Mol. Spectrosc.*, **168**, 271 (1994); A. Császár et al., *J. Chem. Phys.*, **122**, 214305 (2005).

are found by a least-squares fit of the calculated energies. (This formula can be applied using either energies at optimized geometries or energies at a fixed geometry.) There are three unknown quantities in (15.23),  $A$ ,  $B$ , and  $E_{SCF}(\infty)$ , and we have five calculated energies to do the least-squares fit. Although all five energies could be used, experience shows that the accuracy of the extrapolation is usually improved if the energy of the rather small

cc-pVDZ set is not used in the fit. Use of the  $n = 3, 4, 5$ , and 6 energies gives (Prob. 15.21)  $E_{\text{SCF}}(\infty) = -76.0686$  hartrees, close to the  $-76.068_3$  value listed in Table 15.2.

Although the SCF MO calculations give good geometries and fairly good dipole moments, they give poor dissociation energies. For the near-Hartree–Fock calculations of Rosenberg et al., the difference between the energy of two isolated hydrogen atoms and one oxygen atom (as calculated with the oxygen basis set used in the  $\text{H}_2\text{O}$  calculation) and the calculated  $\text{H}_2\text{O}$  energy at the optimized geometry gives  $D_e = 6.94$  eV, compared with the experimental value 10.09 eV. This large error in dissociation energy is typical of Hartree–Fock calculations, as we saw for diatomics. When  $\text{H}_2\text{O}$  is formed from  $2\text{H} + \text{O}$ , two new electron pairs are formed. Two electrons paired in the same MO move through the same region of space, and hence the correlation energy of such a pair is substantial. The Hartree–Fock calculation does not take into account this extra correlation in the molecule (as compared with the separated atoms) and hence gives too small a binding energy.

Hartree–Fock orbital energies have experimental as well as theoretical significance. In 1933, Koopmans gave arguments that indicate that the energy required to remove an electron from a closed-shell atom or molecule is reasonably well approximated by minus the orbital energy  $\varepsilon$  of the AO or MO from which the electron is removed, a result called **Koopmans' theorem**. A partial justification of this result is the fact that, if we neglect the change in the form of the MOs that occurs when the molecule is ionized, then the difference between the Hartree–Fock energies of the ion and the neutral closed-shell molecule can be shown to equal the orbital energy of the MO from which the electron was removed (see Prob. 15.18).

The energy needed to remove an electron from an MO of a molecule can be found experimentally using photoelectron spectroscopy. Here, one uses photons of known energy to knock electrons out of gas-phase molecules and measures the kinetic energies of the emitted electrons. A comparison of minus the near-Hartree–Fock orbital energies (Rosenberg and Shavitt, cited in Table 15.2) with the experimentally observed vertical ionization energies from the various  $\text{H}_2\text{O}$  MOs follows, where the energies are in electronvolts and the experimental values are in parentheses:  $1a_1$ , 559.5 (539.7);  $2a_1$ , 36.7 (32.2);  $1b_2$ , 19.5 (18.5);  $3a_1$ , 15.9 (14.7);  $1b_1$ , 13.8 (12.6). (The *vertical* ionization energy is the energy difference between the molecule  $\text{M}$  and the ion  $\text{M}^+$  when  $\text{M}^+$  is at the equilibrium geometry of  $\text{M}$ , whereas for the *adiabatic* ionization energy,  $\text{M}^+$  is at the equilibrium geometry of  $\text{M}^+$ . In the unlikely event that the equilibrium geometries of  $\text{M}$  and  $\text{M}^+$  are the same, these two ionization energies are equal. Otherwise, the vertical ionization energy is larger than the adiabatic ionization energy. Koopmans' theorem predicts the vertical ionization energy.) Koopmans'-theorem ionization energies are somewhat inaccurate because of (1) neglect of the change in the forms of the MOs that occurs on ionization, and (2) neglect of the change in correlation energy between the neutral molecule and the ion.

## 15.6 Population Analysis and Bond Orders

A widely used (and widely criticized) method to analyze SCF MO wave functions is population analysis, introduced by Mulliken. He proposed a method that apportions the electrons of an  $n$ -electron molecule into net populations  $n_r$  in the basis functions  $\chi_r$  and overlap populations  $n_{r-s}$  for all possible pairs of basis functions.

For the set of basis functions  $\chi_1, \chi_2, \dots, \chi_b$ , each MO  $\phi_i$  has the form

$$\phi_i = \sum_{s=1}^b c_{si} \chi_s = c_{1i} \chi_1 + c_{2i} \chi_2 + \cdots + c_{bi} \chi_b$$

For simplicity, we shall assume that the  $c_{si}$ 's and  $\chi_s$ 's are real. The probability density associated with one electron in  $\phi_i$  is

$$|\phi_i|^2 = c_{1i}^2\chi_1^2 + c_{2i}^2\chi_2^2 + \cdots + 2c_{1i}c_{2i}\chi_1\chi_2 + 2c_{1i}c_{3i}\chi_1\chi_3 + 2c_{2i}c_{3i}\chi_2\chi_3 + \cdots$$

Integrating this equation over three-dimensional space and using the fact that  $\phi_i$  and the  $\chi_s$ 's are normalized, we get

$$1 = c_{1i}^2 + c_{2i}^2 + \cdots + 2c_{1i}c_{2i}S_{12} + 2c_{1i}c_{3i}S_{13} + 2c_{2i}c_{3i}S_{23} + \cdots \quad (15.24)$$

where the  $S$ 's are overlap integrals:  $S_{12} = \int \chi_1\chi_2 dv_1 dv_2$ , etc. Mulliken proposed that the terms in (15.24) be apportioned as follows. One electron in the MO  $\phi_i$  contributes  $c_{1i}^2$  to the net population in  $\chi_1$ ,  $c_{2i}^2$  to the net population in  $\chi_2$ , etc., and contributes  $2c_{1i}c_{2i}S_{12}$  to the overlap population between  $\chi_1$  and  $\chi_2$ ,  $2c_{1i}c_{3i}S_{13}$  to the overlap population between  $\chi_1$  and  $\chi_3$ , etc.

Let there be  $n_i$  electrons in the MO  $\phi_i$  ( $n_i = 0, 1, 2$ ) and let  $n_{r,i}$  and  $n_{r-s,i}$  symbolize the contributions of electrons in the MO  $\phi_i$  to the net population in  $\chi_r$  and to the overlap population between  $\chi_r$  and  $\chi_s$ , respectively. We have

$$n_{r,i} = n_i c_{ri}^2, \quad n_{r-s,i} = n_i (2c_{ri}c_{si}S_{rs}) \quad (15.25)$$

By summing over the occupied MOs  $\phi_i$ , we obtain the Mulliken **net population**  $n_r$  in  $\chi_r$  and the **overlap population**  $n_{r-s}$  for the pair  $\chi_r$  and  $\chi_s$  as

$$n_r = \sum_i n_{r,i} \quad \text{and} \quad n_{r-s} = \sum_i n_{r-s,i}$$

The sum of all the net and overlap populations equals the total number of electrons  $n$  in the molecule (Prob. 15.23):  $\sum_r n_r + \sum_{r>s} \sum_s n_{r-s} = n$ .

### EXAMPLE

For the  $\text{H}_2\text{O}$  MOs in (15.19), calculate the net and overlap population contributions from the  $2a_1$  MO and find  $n_r$  for each basis function. Use  $\text{H}_1 1s$  and  $\text{H}_2 1s$  as basis functions, rather than the symmetry-adapted basis functions.

To find overlap populations, we need the overlap integrals. Since an orthogonalized  $2s$  AO is used, all the basis functions centered on oxygen are mutually orthogonal, and the overlap populations are zero for all pairs of functions centered on oxygen. (For extended basis sets, this is not true. For example, a DZ basis set uses two  $1s$ -type functions on oxygen, and these functions are not orthogonal to each other.) Overlap integrals between basis STOs centered on different atoms can be found by interpolation in the tables of R. S. Mulliken et al., *J. Chem. Phys.*, **17**, 1248 (1949). One finds (Prob. 15.24)

$$\begin{aligned} \langle \text{H}_1 1s | \text{O} 1s \rangle &= \langle \text{H}_2 1s | \text{O} 1s \rangle = 0.054, & \langle \text{H}_1 1s | \text{O} 2s_{\perp} \rangle &= \langle \text{H}_2 1s | \text{O} 2s_{\perp} \rangle = 0.471 \\ \langle \text{H}_1 1s | \text{O} 2p_y \rangle &= -\langle \text{H}_2 1s | \text{O} 2p_y \rangle = 0.319, & \langle \text{H}_1 1s | \text{O} 2p_z \rangle &= \langle \text{H}_2 1s | \text{O} 2p_z \rangle = 0.247 \\ \langle \text{H}_1 1s | \text{H}_2 1s \rangle &= 0.238 \end{aligned}$$

The contributions to the net populations in the basis functions from the two electrons in the  $2a_1$  MO are given by (15.25) and (15.19) as

$$\begin{aligned} n_{\text{O}1s,2a_1} &= 2(-0.027)^2 = 0.0015, & n_{\text{O}2s_{\perp},2a_1} &= 2(0.820)^2 = 1.345, & n_{\text{O}2p_y,2a_1} &= 0.035 \\ n_{\text{H}_1 1s,2a_1} &= 2(0.152)^2 = 0.046, & n_{\text{H}_2 1s,2a_1} &= 0.046 \end{aligned}$$

The  $2a_1$  contributions to the nonzero overlap populations are given by (15.25) and (15.19) as

$$n_{O1s-H1s,2a_1} = 2(2)(-0.027)(0.152)(0.054) = -0.0009 = n_{O1s-H2s,2a_1}$$

$$n_{O2s_{\perp}-H1s,2a_1} = 0.235 = n_{O2s_{\perp}-H2s,2a_1}$$

$$n_{O2p_z-H1s,2a_1} = 0.020 = n_{O2p_z-H2s,2a_1}, \quad n_{H1s-H2s,2a_1} = 0.022$$

The net population of  $O1s$  is found from (15.25) and (15.19) as the sum of contributions from each occupied MO:

$$n_{O1s} = 2(1.000)^2 + 2(-0.027)^2 + 2(-0.026)^2 = 2.00$$

The net populations for the other basis functions are (Prob. 15.25a)  $n_{O2s_{\perp}} = 1.85$ ,  $n_{O2p_x} = 2.00$ ,  $n_{O2p_y} = 0.78$ ,  $n_{O2p_z} = 1.27$ ,  $n_{H1s} = 0.545$ ,  $n_{H2s} = 0.545$ .

To decide whether the MO  $\phi_i$  in a covalent molecule is bonding, we examine the sum of those overlap-population contributions  $n_{r-s,i}$  for which the basis functions  $\chi_r$  and  $\chi_s$  lie on different atoms. If this interatomic overlap population contribution is substantially positive, the MO is bonding; if it is substantially negative, the MO is antibonding. If it is zero or near zero, the MO is nonbonding.

For example, for the  $3a_1$  MO in (15.19), overlap of  $O1s$  with  $H1s$  contributes  $2(2)(-0.026)(0.264)(0.054) = -0.0015$ , overlap of  $O1s$  with  $H2s$  contributes  $-0.0015$ , overlap of  $O2s_{\perp}$  with  $H1s$  contributes  $2(2)(-0.502)(0.264)(0.471) = -0.250$ , overlap of  $O2s_{\perp}$  with  $H2s$  contributes  $-0.250$ , overlap of  $O2p_z$  with  $H1s$  contributes  $0.205$  and with  $H2s$  contributes  $0.205$ , and overlap of  $H1s$  with  $H2s$  contributes  $2(2)(0.264)^2(0.238) = 0.066$ . Summing, we get an interatomic overlap population of  $-0.03$  for the  $3a_1$  MO. This value is near zero, indicating a nonbonding (lone-pair) MO. The interatomic overlap population for  $2a_1$  is found to be  $0.53$ , and that for  $1b_2$  is  $0.50$  (Prob. 15.25). These are bonding MOs. For the inner-shell  $1a_1$  MO, we get  $0.00$ .

Instead of apportioning the electrons into net populations in basis functions and overlap populations for pairs of basis functions, it is convenient for some purposes to apportion the electrons among the basis functions only, with no overlap populations. Mulliken proposed that this be done by splitting each overlap population  $n_{r-s}$  equally between the basis functions  $\chi_r$  and  $\chi_s$ . For each basis function  $\chi_r$ , this gives a **gross population**  $N_r$  in  $\chi_r$  that equals the net population  $n_r$  plus one-half the sum of the overlap populations between  $\chi_r$  and all other basis functions:

$$N_r = n_r + \frac{1}{2} \sum_{s \neq r} n_{r-s}$$

The sum of all the gross populations equals the number of electrons in the molecule:  $\sum_{r=1}^b N_r = n$ .

For example, the contribution to the gross population of  $O2s_{\perp}$  from the  $2a_1$  MO in (15.19) is

$$N_{O2s_{\perp},2a_1} = 2[(0.820)^2 + (0.820)(0.152)(0.471) + (0.820)(0.152)(0.471)] = 1.58$$

One finds these other contributions to the gross population of  $O2s_{\perp}$ :  $0.00$  from  $1a_1$ ,  $0.25$  from  $3a_1$ , zero from  $1b_2$  and  $1b_1$ . Addition of these contributions gives a gross population (or occupation number) of  $1.83$  for  $O2s_{\perp}$ . Doing the calculation for the other basis functions, one finds (Prob. 15.25c) these gross populations (where the contributions are listed in the order  $1a_1, 2a_1, 1b_2, 3a_1, 1b_1$ ):  $N_{O1s} = 2.00 + 0.00 + 0 + 0.00 + 0 = 2.00$ ;  $N_{O2s_{\perp}} = 1.83$ ;

$$N_{O2p_x} = 0 + 0 + 0 + 0 + 2 = 2; \quad N_{O2p_y} = 0 + 0 + 1.12 + 0 + 0 = 1.12; \quad N_{O2p_z} = 0 + 0.05_5 + 0 + 1.44_5 + 0 = 1.50; \quad N_{H1s} = 0.00 + 0.18_4 + 0.44_2 + 0.150 + 0 = 0.77_6; \quad N_{H21s} = 0.77_6.$$

Addition of the gross populations for all basis functions centered on atom B gives the **gross atomic population**  $N_B$  for atom B:

$$N_B = \sum_{r \in B} N_r$$

where the notation  $r \in B$  denotes all basis functions centered on atom B. Provided all basis functions are atom centered (this is usually true), the sum of the gross atomic populations equals the number of electrons in the molecule. The Mulliken **net atomic charge**  $q_B$  on atom B with atomic number  $Z_B$  is defined as

$$q_B \equiv Z_B - N_B$$

For example, for the Pitzer–Merrifield  $H_2O$  wave function, the gross atomic populations, found by summing the gross populations of the basis functions on each atom, are  $N_O = 2.00 + 1.83 + 2 + 1.12 + 1.50 = 8.45$ ,  $N_{H_1} = 0.77_6 = N_{H_2}$ ; the net charges are  $q_O = 8 - 8.45 = -0.45$ ,  $q_{H_1} = 1 - 0.77_6 = 0.22_4 = q_{H_2}$ . As expected, the oxygen is negatively charged.

One should not rely too much on numbers calculated by population analysis. Mulliken's assignment of half the overlap population to each basis function is arbitrary and sometimes leads to unphysical results (see *Mulliken and Ermler, Diatomic Molecules*, pages 36–38, 88–89). Moreover, a small change in basis set can produce a large change in the calculated net charges. For example, net atomic charges on each H atom in  $CH_4$ ,  $NH_3$ ,  $H_2O$ , and HF calculated from HF/STO-3G, HF/3-21G, HF/6-31G\*, and HF/6-31G\*\* wave functions are (CCCBDB, Section 15.1)

|         | CH <sub>4</sub> | NH <sub>3</sub> | H <sub>2</sub> O | HF   |
|---------|-----------------|-----------------|------------------|------|
| STO-3G  | 0.07            | 0.15            | 0.17             | 0.19 |
| 3-21G   | 0.20            | 0.29            | 0.37             | 0.45 |
| 6-31G*  | 0.16            | 0.33            | 0.43             | 0.52 |
| 6-31G** | 0.12            | 0.26            | 0.34             | 0.39 |

Comparison of values calculated with the same basis set correctly shows the increasing charge on each H atom as the electronegativity increases from C to N to O to F, but comparison of values calculated with different basis sets could erroneously lead one to say that the C—H bond in  $CH_4$  is more polar than the F—H bond in FH.

An improvement on Mulliken population analysis (MPA) is **natural population analysis** (NPA) [A. E. Reed, R. B. Weinstock, and F. Weinhold, *J. Chem. Phys.*, **83**, 735 (1985)], which uses ideas related to natural orbitals (*Pilar*, Section 10-7). Here, one first calculates a set of orthonormal natural atomic orbitals (NAOs) from the AO basis set  $\chi_r$ . The NAOs are then used to compute a set of orthonormal **natural bond orbitals** (NBOs), where each occupied NBO is classifiable as a core, lone pair, or bond orbital. Using these NBOs, one carries out a population analysis. NPA net atomic charges show less basis-set dependence than those from Mulliken population analysis. Other methods of assigning net atomic charges are discussed in the next section. Still another method of population analysis that yields net atomic charges is *Löwdin population analysis* (*Cramer*, Section 9.1.3.2).

A review of population analysis recommended that in view of the existence of improved methods, Mulliken population analysis should no longer be used [S. M. Bachrach in K. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry*, vol. 5, VCH (1994), Chapter 3.]

### Bond Orders

Many methods have been proposed to calculate bond orders from SCF MO wave functions. One useful definition is due to Mayer [I. Mayer, *Chem. Phys. Lett.*, **97**, 270 (1983)]. For a closed-shell molecule whose occupied MOs are  $\phi_i = \sum_s c_{si} \chi_s$ , the Mayer bond order  $B_{CD}$  between atoms C and D is defined as

$$B_{CD} \equiv \sum_{t \in C} \sum_{u \in D} \left( \sum_r P_{tr} S_{ru} \right) \left( \sum_r P_{ur} S_{rt} \right) \quad (15.26)$$

where the density matrix element  $P_{tr}$  is given by (14.42) as  $P_{tr} \equiv \sum_j n_j c_{tj}^* c_{uj}$ , the notation  $t \in C$  denotes all basis functions centered on atom C,  $S_{rt}$  is the overlap integral between basis functions  $\chi_r$  and  $\chi_t$ , and  $n_j$  is the number of electrons in MO  $j$ . The Mayer bond orders usually have values close to those found from Lewis electron-dot formulas.

## 15.7 The Molecular Electrostatic Potential, Molecular Surfaces, and Atomic Charges

### The Molecular Electrostatic Potential

The **electric potential**  $\phi$  at a point P in space is defined as the reversible work per unit charge needed to move an infinitesimal test charge  $Q_t$  from infinity to P, which we write as  $\phi_P \equiv w_{\infty \rightarrow P}/Q_t$ . The SI unit of  $\phi$  is the volt (V), where  $1 \text{ V} \equiv 1 \text{ J/C}$ . When we do reversible work  $w$  on the test charge, we change its potential energy  $V$  by  $w$  (just as reversibly raising or lowering a mass in the earth's gravitational field changes its potential energy). If we take the potential energy of  $Q_t$  as zero at infinity, we therefore have  $V_P = w_{\infty \rightarrow P} = \phi_P Q_t$ . The electrical potential energy  $V$  of a charge at point P (where the electric potential is  $\phi_P$ ) is  $\phi_P Q_t$ . From the definition of  $\phi_P$ , it readily follows (Prob. 15.26) that in the space around a point charge  $Q$ , the electric potential (in SI units) is  $\phi_P = Q/4\pi\epsilon_0 d$ , where  $d$  is the distance between point P and the charge. The electric potential is a function of the location  $(x, y, z)$  of point P in space:  $\phi = \phi(x, y, z)$ .

If the system consists of a single point charge  $Q_A$  located at  $(x_A, y_A, z_A)$ , then  $\phi_1 = Q_A/4\pi\epsilon_0 r_{1A}$ , where  $r_{1A}$  is the distance between point A and point 1 with coordinates  $(x_1, y_1, z_1)$ . If the system consists of several point charges, then each contributes to  $\phi$  and

$$\phi_1 = \sum_i \frac{Q_i}{4\pi\epsilon_0 r_{1i}} \quad (15.27)$$

If the system is a molecule, we view it as a collection of point-charge nuclei and electronic charge smeared out into a continuous distribution. Electrons are point charges and are not actually smeared out into a continuous charge distribution, but the electronic-charge-cloud picture is a reasonable approximation when considering interactions between two molecules that are not too close to each other. The probability of finding a molecular electron in a tiny volume  $dv = dx dy dz$  is  $\rho dv$ , where  $\rho$  is the electron probability density [Eq. (14.5)]. Therefore, the amount of electronic charge in  $dv$  is  $-e\rho dv$ . Addition of the contributions of the molecular electronic charge and of the nuclei  $\alpha$  gives the molecular electric potential as

$$\phi(x_1, y_1, z_1) = \sum_{\alpha} \frac{Z_{\alpha} e}{4\pi\epsilon_0 r_{1\alpha}} - e \iiint \frac{\rho(x_2, y_2, z_2)}{4\pi\epsilon_0 r_{12}} dx_2 dy_2 dz_2 \quad (15.28)$$

where  $r_{12}$  is the distance between points 1 and 2 and the integration is over all space. In atomic units, the  $e$  and the  $4\pi\epsilon_0$  disappear from (15.28).

Quantum chemists call  $\phi$  the **molecular electrostatic potential** (MEP) or the **electrostatic potential** (ESP—no connection with parapsychology). Although the SI units of  $\phi$  are volts, quantum chemists traditionally multiply (15.28) by the proton charge  $e$  (thereby converting its units to joules) and by the Avogadro constant  $N_A$  (thereby converting its units to J/mol). Thus a quantum chemist's MEP value at a point P is the molar electrical interaction energy between the molecule and a proton placed at point P, assuming that the molecule is not polarized by the proton.

To calculate  $\phi(x, y, z)$  one calculates an approximate electronic wave function for the equilibrium geometry (Section 15.10), uses (14.5) to calculate  $\rho(x, y, z)$  from the approximate wave function, and uses (15.28) to calculate  $\phi$ . [In density-functional theory (DFT) (Section 16.5),  $\rho$  is calculated without first finding the molecular wave function.] The MEP is found not to be strongly affected by the choice of basis set or by the inclusion of electron correlation. The MEP is often calculated at the HF/6-31G\* level of theory or using DFT. The MEP can be calculated from accurate experimental X-ray-diffraction crystallography data [N. Bouhmadia et al., *J. Chem. Phys.*, **116**, 6196 (2002)].

The electron probability density  $\rho(x, y, z)$  tells us how the electronic charge is distributed in a molecule. The MEP  $\phi(x, y, z)$  tells us the interaction energy between a non-polarizing positive test charge at  $(x, y, z)$  and the nuclear charges and electronic charge distribution of the molecule. The MEP is commonly depicted by a contour map showing curves of constant  $\phi$  in a particular plane through the molecule or by a surface in three-dimensional space on which  $\phi$  is constant. MEPs are generally found to be positive throughout the spatial region within a molecule, due to the strong positive contributions from the nuclei. Outside a molecule,  $\phi$  can be positive or negative. For each of the molecules  $\text{H}_2\text{O}$  and  $\text{HC}(\text{O})\text{NH}_2$ , the MEP is negative in the region outside the oxygen atom and positive outside other atoms.

When two molecules approach each other, the MEP of each plays a key role in their interaction. An electrophilic (electron-loving) species will preferentially attack a molecule at sites where the MEP is most negative. (This is not invariably true, since the MEP ignores polarization of the molecule by the incoming species.) MEPs provide insight into molecular recognition processes such as enzyme–substrate and drug–receptor interactions.

If we model each atom in a molecule as a sphere of radius equal to the van der Waals radius of the atom (for bonded atoms, these spheres overlap, and the spheres of bonded atoms are truncated), the **van der Waals surface** of a molecule is defined by the outward-facing surfaces of these atomic spheres. The van der Waals surface is what one sees in the familiar space-filling CPK (R. B. Corey–Pauling–Koltun) molecular models. In discussing intermolecular interactions, the MEPs in the regions on and outside the van der Waals surface are most significant.

A more sophisticated surface than the van der Waals surface is an **isodensity surface**, defined as a surface on which the molecule's electron probability density  $\rho$  is constant. The molecular isodensity surface for which  $\rho = 0.001 \text{ electrons/bohr}^3 = 0.006748 \text{ electrons/\AA}^3$  was found to have about the same surface area as the van der Waals surface [C. K. Kim et al., *J. Comput. Chem.*, **25**, 2073 (2004)], and this isodensity surface is often used, although many workers prefer the  $0.002 \text{ electrons/bohr}^3$  isodensity surface. For certain purposes, one wants values of the **molecular surface area** and the **molecular volume**. Of course, these quantities have no well-defined meaning, but these values are usually calculated as the area of and the volume enclosed by the van der Waals surface or by an isodensity surface with a specified  $\rho$  value.

As an alternative to showing surfaces of constant MEP, one often examines the MEP values on an isodensity surface of the molecule. These values constitute the **molecular surface electrostatic potential** (MSEP or MESP). The MSEP is often depicted using a

color scheme such as red denoting the most negative MEP values, blue the most positive, and other spectrum colors intermediate values. Negative regions of the MSEP arise from lone pairs on electronegative atoms, from  $\pi$  electrons (for example, the MSEP of benzene is negative within much of the hexagon above and below the molecular plane), and from strained carbon–carbon bonds (the MSEP of cyclopropane is negative in the three regions that lie near to and outside the three carbon–carbon bonds and near the plane of the C atoms). To see many examples of MSEP's, click on Images in Google and enter molecular surface electrostatic potential. (You can also create such surfaces using some quantum chemistry programs; see Prob. 15.49.)

Politzer and co-workers defined several statistical quantities (such as the average MSEP, the standard deviation of the MSEP) that are calculated from the MSEP and found empirical relations between condensed-phase organic-compound macroscopic properties that depend on intermolecular interactions (such as normal boiling point, heats of fusion and vaporization, surface tension) and the MSEP quantities. This allows prediction of these macroscopic properties from the MSEP with a fair degree of success [P. Politzer and J. S. Murray, *Fluid Phase Equilib.*, **185**, 129 (2001)].

For information on the role of the MEP in biochemical processes, see B. Honig and A. Nicholls, *Science*, **268**, 1144 (1995). For reviews of MEPs, see P. Politzer and J. S. Murray in K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry*, vol. 2., VCH, 1991, Chapter 7; P. Politzer and J. S. Murray, *Theor. Chem. Acc.*, **108**, 134 (2002); Murray and Politzer, *WIREs Comput. Mol. Sci.*, **1**, 153 (2011).

### Atomic Charges

Whereas the MEP is a well-defined, physically significant quantity that can be accurately calculated from a reasonably accurate wave function, there is no unique, well-defined answer to the question: What is the charge on a particular atom in a molecule? (The terms *atomic charge*, *net atomic charge*, and *partial atomic charge* are used synonymously.)

As noted in Section 15.6, Mulliken population analysis gives atomic charges that vary erratically as the basis set is improved. Better values are obtained from natural population analysis (NPA).

A popular way to get reasonable atomic charges  $Q_\alpha$  is by fitting the MEP  $\phi$ . One first uses a molecular wave function to calculate values of  $\phi$  at a grid of many points in the region outside the molecule's van der Waals surface. One then places a charge  $Q_\alpha$  at each nucleus  $\alpha$  and calculates the quantity  $\phi_i^{\text{approx}} \equiv \sum_\alpha Q_\alpha e / 4\pi\epsilon_0 r_{i\alpha}$  at each grid point. One then varies the  $Q_\alpha$  values (subject to the constraint that they add to zero for a neutral molecule) so as to minimize the sum of the squares of the deviations  $\phi_i^{\text{approx}} - \phi_i$  for the grid points. Various ways of choosing the grid points and of including other refinements give different schemes for finding the atomic charges  $Q_\alpha$ , which are called **ESP** (electrostatic potential) **charges**. Three common schemes are the Merz–Singh–Kollman (MK or MSK) method, the CHELPG method (charges from electrostatic potentials, grid method), and the RESP (restrained ESP) method. The MEP depends most strongly on the charges assigned to atoms near the surface of the molecule and depends only quite weakly on the charges of atoms in the interior, so ESP charges of atoms in the interior cannot be accurately assigned.

Another procedure that yields atomic charges is Bader's **atoms-in-molecules** (AIM) theory [Bader; R. F. W. Bader, *Chem. Rev.*, **91**, 893 (1991)]. Here, one first calculates an approximate wave function for the molecule, and from this wave function one calculates the electron probability density  $\rho(x, y, z)$ . One now imagines drawing all the contour surfaces of constant  $\rho$  (isodensity surfaces). Then one draws lines (called *gradient paths*) throughout three-dimensional space such that at every point on a given line, that line is perpendicular to the isodensity surface that passes through the point. (A simple example is

the ground-state H atom. Here the isodensity surfaces are spheres centered at the nucleus, and the gradient paths are radii emanating from the nucleus.) One finds that most such gradient paths start at infinity and end at one of the nuclei. For molecules containing rings (for example, benzene), many gradient paths will go from the point within the ring where  $\rho$  is a minimum to a nucleus. The three-dimensional region of space  $\Omega_A$  belonging to atom A of the molecule is then defined as the region that contains all the gradient paths that end at the nucleus of atom A. The AIM charge on atom A with atomic number  $Z_A$  is then defined by  $Q_A = Z_A - \int_{\Omega_A} \rho dV$ , where the integration goes over the region  $\Omega_A$ .

The gradient vector  $\nabla\rho$  at a point P can be shown to be perpendicular to the isodensity surface  $\rho = \text{constant}$  at P. There is only one perpendicular direction to a surface at a particular point, so a line drawn perpendicular to isodensity surfaces will have its direction at any point be in the direction of  $\nabla\rho$ . Hence such a line is called a *gradient path*. Because  $\nabla\rho$  has a unique direction at each point in space (except for points where  $\nabla\rho$  is zero or undefined), gradient paths from different nuclei cannot cross each other, and the gradient paths terminating at each nucleus divide space into nonoverlapping regions, one for each atom.

AIM charges have been criticized for being larger than seems chemically reasonable and, according to Cramer, “are of little chemical utility” (Cramer, p. 317). For Bader’s defense of AIM charges, see R. F. W. Bader and C. F. Matta, *J. Phys. Chem. A*, **108**, 8385 (2004). [See also C. F. Matta and R. F. W. Bader, *J. Phys. Chem. A*, **110**, 6365 (2006).]

Cramer, Truhlar, and co-workers have devised the methods **Charge-Model 1** (CM1), CM2, CM3, CM4, and CM5 to calculate atomic charges [Cramer, Sec. 9.1.3.4; C. P. Kelly et al., *Theor. Chem. Acc.*, **113**, 133 (2005); comp.chem.umn.edu/solvation/comparison.htm#charges]. These methods start with a set of charges such as the Mulliken charges or the Löwdin charges and then improve these charges by using a specified formula that transfers charges from one atom to another to get the final CM charges. The formula used to transfer charges contains parameters whose values have been fixed at those values that allow CM-charge-calculated molecular dipole moments  $\mu^{\text{CM}}$  of a large set of training molecules to give the best fit to the experimental dipole moments of these molecules, where the CM dipole moments are calculated from  $\mu^{\text{CM}} = |\sum_B Q_B^{\text{CM}} \mathbf{r}_B|$ , where  $Q_B^{\text{CM}}$  is the CM charge on atom B, and the sum goes over the atoms of the molecule. More specifically, CM2 and CM3 use the formula

$$Q_k^{\text{CM}} = Q_k^{\text{L}} + \sum_{k' \neq k} B_{kk'}^{\text{M}} (D_{kk'} + C_{kk'} B_{kk'}^{\text{M}})$$

where  $Q_k^{\text{L}}$  is the charge given by Löwdin population analysis,  $B_{kk'}^{\text{M}}$  is the Mayer bond order for atoms  $k$  and  $k'$ , the sum goes over all atoms except atom  $k$ , and the values of the parameters  $C_{kk'}$  and  $D_{kk'}$  depend only on the atomic numbers of the two atoms. Different sets of  $C$  and  $D$  parameters are used depending on which method (HF, DFT, etc.) and which basis set are used to find  $Q_k^{\text{L}}$ .

The Charge Model 5 (CM5) method [A. V. Marenich et al., *J. Chem. Theory Comput.*, **8**, 527 (2012)] of finding atomic charges uses the formula

$$Q_k^{\text{CM5}} = Q_k^{\text{H}} + \sum_{k' \neq k} T_{kk'} B_{kk'}$$

where  $Q_k^{\text{H}}$  is the Hirshfeld charge (discussed in the next paragraph) on atom  $k$ ; the sum goes over all atoms in the molecule except  $k$ ;  $T_{kk'}$  is a parameter that depends on which elements atoms  $k$  and  $k'$  are; and  $B_{kk'} \equiv \exp[-\alpha(r_{kk'} - R_k - R_{k'})]$ , where  $\alpha$  is a parameter,  $r_{kk'}$  is the distance between atoms  $k$  and  $k'$  in the molecule, and  $R_k$  and  $R_{k'}$  are the tabulated atomic covalent radii of atoms  $k$  and  $k'$ . The values of the parameters were chosen so that the method gave a good fit to the dipole moments of a large series of molecules.

Another method to find atomic charges is the Hirshfeld method. Here, atomic charges are obtained by distributing the molecular electronic probability density at each point in space among the atoms in the molecule according to the formula

$$\rho_A(x, y, z) = \frac{\rho_A^0(x, y, z)}{\sum_A \rho_A^0(x, y, z)} \rho_{\text{mol}}(x, y, z)$$

where  $\rho_{\text{mol}}(x, y, z)$  is the molecular electronic probability density at a point (calculated from some approximate wave function),  $\rho_A(x, y, z)$  is the portion of the electronic density at  $(x, y, z)$  being assigned to atom A,  $\rho_A^0(x, y, z)$  is the electronic density computed for the isolated atom A (calculated using the same method and basis set used to get  $\rho_{\text{mol}}$ ), and the sum is evaluated by adding up the probability densities of isolated atoms each located at where it is in the molecule. The charge  $Q_A$  on atom A is then found as  $Q_A = Z_A - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_A dx dy dz$ , where  $Z_A$  is the atomic number of A. Because Hirshfeld charges have several deficiencies, the improved **Hirshfeld-I** method (where I stands for iterative) was proposed [P. Bultnick et al., *J. Chem. Phys.*, **126**, 144111 (2007)]. This method starts from the Hirshfeld charges and uses an iterative process that converges to a set of improved charges. A modification of the Hirshfeld-I method is the *valence Hirshfeld-I method*, which divides only the valence electronic charge density among the atoms, and assigns all the density of the core electrons of an atom to that atom [L. Vanduyfhuys et al., *J. Chem. Theory Comput.*, **8**, 3217 (2012)].

An assessment of several atomic charge methods applied to gas-phase polypeptide conformers found that the Hirshfeld-I method performed best overall in reproducing MP2-calculated electric dipole moments and in being transferable from one conformer of a molecule to another [T. Verstraelen et al., *J. Chem. Theory Comput.*, **8**, 661 (2012)]. Also, Hirshfeld-I charges reproduce the molecular electrostatic potential with good accuracy.

Calculated atomic charges are used as parameters in molecular-mechanics calculations (Section 17.5) to model the electrostatic interactions between nonbonded atoms. ESP-derived charges and CM1 charges are often used for this purpose.

For reviews of atomic charges, see M. M. Francl and L. E. Chirlian in K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry*, vol. 14, Wiley, 2000, Chapter 1; *Cramer*, Section 9.1.3.

## 15.8 Localized MOs

The idea of a chemical *bond* between a pair of atoms in a molecule is fundamental to chemistry. The experimental evidence supporting this concept is substantial. One can assign bond energies to various kinds of bonds and obtain good estimates for the heats of atomization of most molecules by adding the energies of the individual bonds. Other molecular properties (for example, magnetic susceptibility, dipole moment) can also be analyzed as the sum of contributions from individual bonds (and lone pairs). The infrared spectrum of a compound containing an OH group shows a characteristic band near  $3600 \text{ cm}^{-1}$ . This OH stretching vibrational band occurs at nearly the same frequency in HOH, HOCl, and  $\text{CH}_3\text{OH}$ . The length of an OH bond is nearly constant from molecule to molecule, about  $0.96 \text{ \AA}$ .

The MO picture of  $\text{H}_2\text{O}$  presented in Section 15.5 appears to be gravely deficient, in that it is seemingly inconsistent with the existence of individual bonds in the molecule. Each of the bonding MOs in (15.19) is delocalized over the entire molecule. If one were to compare the bonding MOs for HOH and HOCl, one would find them to be quite different. Yet we know that the OH parts of these two molecules are similar. Actually, MO theory *can* explain the observed near invariance of a given kind of chemical bond, as we now show.

The MO approximation to the ground state of H<sub>2</sub>O is a Slater determinant of the form

$$|\phi_1\bar{\phi}_1\phi_2\bar{\phi}_2\phi_3\bar{\phi}_3\phi_4\bar{\phi}_4\phi_5\bar{\phi}_5| \quad (15.29)$$

Putting (15.29) into words, we say that two electrons are in each of the orthonormal MOs  $\phi_1, \phi_2, \phi_3, \phi_4$ , and  $\phi_5$ . However, this MO description is not unique. The addition of a multiple of one column of a determinant to another column leaves the determinant unchanged in value. Adding column 7 to column 9 and column 8 to column 10 in (15.29), we get

$$|\phi_1\bar{\phi}_1\phi_2\bar{\phi}_2\phi_3\bar{\phi}_3\phi_4\bar{\phi}_4(\phi_4 + \phi_5)(\bar{\phi}_4 + \bar{\phi}_5)| \quad (15.30)$$

This determinant leads to the description of the molecule as having the MOs  $\phi_1, \phi_2, \phi_3, \phi_4$ , and  $\phi_4 + \phi_5$  each doubly occupied. Despite the different verbal descriptions, the wave functions (15.29) and (15.30) are identical.

[In discussing SCF calculations, we used an orthogonalized  $2s$  AO of the form  $a(1s) + b(2s)$ , where  $2s$  is a (nodeless)  $2s$  Slater-type orbital. This procedure is justified by use of the freedom of adding a multiple of one column of a determinant to another. The determinantal wave function for the oxygen atom is the same whether it is the  $2s$  or the orthogonalized  $2s_{\perp}$  AO that is used.]

An objection that can be raised to (15.30) is that the MO  $\phi_4 + \phi_5$  is neither normalized nor orthogonal to  $\phi_4$ . Consider, however, the Slater determinant

$$|\phi_1\bar{\phi}_1\phi_2\bar{\phi}_2\phi_3\bar{\phi}_3(b\phi_4 + c\phi_5)(\bar{b}\phi_4 + \bar{c}\phi_5)(c\phi_4 - b\phi_5)(\bar{c}\phi_4 - \bar{b}\phi_5)| \quad (15.31)$$

where  $b$  and  $c$  are any two real constants such that

$$b^2 + c^2 = 1 \quad (15.32)$$

We now show (15.31) and (15.29) to be the same wave function. Multiplication of columns 7 and 8 of (15.29) by  $b$  and columns 9 and 10 by  $-b^{-1}$  gives

$$|\dots b\phi_4(\bar{b}\phi_4)(-b^{-1}\phi_5)(\bar{-b^{-1}\phi_5})|$$

This step multiplies (15.29) by  $b^2(-b^{-1})^2 = 1$ . Next,  $-bc$  times column 9 is added to column 7 and  $-bc$  times column 10 is added to column 8 to give

$$|\dots (b\phi_4 + c\phi_5)(\bar{b}\phi_4 + \bar{c}\phi_5)(-b^{-1}\phi_5)(\bar{-b^{-1}\phi_5})|$$

Finally,  $c/b$  times column 7 is added to column 9,  $c/b$  times column 8 is added to column 10, and (15.32) is used. This gives

$$|\dots (b\phi_4 + c\phi_5)(\bar{b}\phi_4 + \bar{c}\phi_5)(c\phi_4 - b\phi_5)(\bar{c}\phi_4 - \bar{b}\phi_5)|$$

This completes the proof. The orthonormality of the MOs  $c\phi_4 - b\phi_5$  and  $b\phi_4 + c\phi_5$  readily follows from (15.32) and the orthonormality of  $\phi_4$  and  $\phi_5$ . Thus we can describe the molecule as having the orthonormal MOs

$$\phi_1, \phi_2, \phi_3, b\phi_4 + c\phi_5, c\phi_4 - b\phi_5 \quad (15.33)$$

each doubly occupied.

There are an infinite number of other MO descriptions consistent with (15.29): Let  $d$  and  $e$  be any two real constants such that  $d^2 + e^2 = 1$ . If the procedure used to go from (15.29) to (15.31) is applied to columns 5, 6, 7, and 8 of (15.31), we end up with

$$|\phi_1\bar{\phi}_1\phi_2\bar{\phi}_2(d\phi_3 + be\phi_4 + ce\phi_5)(\bar{d}\phi_3 + \bar{b}e\phi_4 + \bar{c}e\phi_5) \\ \times (e\phi_3 - bd\phi_4 - cd\phi_5)(\bar{e}\phi_3 - \bar{b}d\phi_4 - \bar{c}d\phi_5)(c\phi_4 - b\phi_5)(\bar{c}\phi_4 - \bar{b}\phi_5)|$$

We can thus describe the electronic configuration as having the orthonormal MOs

$$\phi_1, \phi_2, (d\phi_3 + be\phi_4 + ce\phi_5), (e\phi_3 - bd\phi_4 - cd\phi_5), (c\phi_4 - b\phi_5)$$

each doubly occupied. Continuing on in this manner, we can derive orthonormal MOs, each of which is a linear combination of all the original MOs  $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5$ . For the general conditions that must be satisfied by the coefficients in these linear combinations, see *Szabo and Ostlund*, Section 3.2.3.

Thus for a given closed-shell electronic state of a molecule, there are many possible MO descriptions. The delocalized MOs (15.19) of  $\text{H}_2\text{O}$  are uniquely determined as the solutions of the SCF equations [Eq. (14.25)]

$$\hat{F}\phi_i = \varepsilon_i\phi_i, \quad i = 1, 2, \dots \quad (15.34)$$

Hence you may be wondering how we can have other sets of MOs that also minimize the variational integral. Actually, Eq. (15.34) is not the most general equation satisfied by the MOs that minimize the variational integral. Instead, one finds in deriving the SCF equations that the Hartree–Fock MOs must satisfy

$$\hat{F}\phi_i = \sum_{j=1}^N \lambda_{ji}\phi_j, \quad i = 1, \dots, N \quad (15.35)$$

where the sum runs over the occupied MOs. The  $\lambda_{ji}$ 's are certain constants (Lagrangian multipliers), which can be chosen arbitrarily, subject to the MO orthonormality requirement. Different choices of the  $\lambda_{ji}$ 's give different sets of MOs, but each such set minimizes the variational integral. For a closed-shell configuration, one possible choice of the  $\lambda_{ji}$ 's can be shown to be

$$\lambda_{ji} = \varepsilon_i\delta_{ij} \quad (15.36)$$

This choice reduces (15.35) to (15.34). [For the proof of (15.35) and (15.36), see *Szabo and Ostlund*, Chapter 3.]

The delocalized MOs [such as (15.19) for  $\text{H}_2\text{O}$ ] that satisfy (15.34) are called the **canonical MOs**. The canonical MOs are eigenfunctions of the Fock operator, which commutes with the molecular symmetry operators, and hence the canonical MOs each transform according to one of the possible molecular symmetry species. A set of MOs like (15.33) satisfies (15.35) but not (15.34) and does not necessarily transform according to the molecular symmetry species.

For an open-shell configuration, one can also find canonical delocalized MOs that satisfy (15.34) and that transform according to the molecular symmetry species. However, the form of  $\hat{F}$  is more complicated than that for the closed-shell case. [See C. C. J. Roothaan, *Rev. Mod. Phys.*, **32**, 179 (1960).]

Of the possible MO sets formed as linear combinations of the delocalized canonical MOs of water, a set that would appeal to a chemist is one for which the charge probability density of each bonding MO is localized in the region of one of the OH bonds. There are many ways of taking linear combinations of the delocalized MOs to get such **localized MOs (LMOs)**. A natural requirement that the two localized bonding MOs in water should satisfy is that they be equivalent to each other; that is, each localized bonding MO of water should be transformed into the other by the  $\hat{C}_2(z)$  symmetry operation (Fig. 15.7). Localized MOs that are permuted among one another by a symmetry operation that permutes equivalent chemical bonds are called **equivalent orbitals**. As we shall see, localized MOs reconcile MO theory with the chemist's intuitive picture of chemical bonding.

For  $\text{H}_2\text{O}$ , the Lewis electron-dot formula suggests the localized MOs to be a pair of equivalent bonding orbitals  $b(\text{OH}_1)$  and  $b(\text{OH}_2)$ , an inner-shell  $1s$  oxygen orbital  $i(\text{O})$ , and two lone-pair equivalent MOs  $l_1(\text{O})$  and  $l_2(\text{O})$  on oxygen.

The most “localized” MOs are those that are most separated from one another. By this, we mean that set of MOs for which the interelectronic repulsion between the different MOs viewed as “charge clouds” is a minimum. The energy of repulsion between the charge

clouds of electron 1 in the MO  $\phi_i$  and electron 2 in the MO  $\phi_j$  is a Coulomb integral of the form (9.99). The total interorbital charge-cloud repulsion energy is then

$$4 \sum_i \sum_{j>i} \iint |\phi_i(1)|^2 |\phi_j(2)|^2 \frac{1}{r_{12}} dv_1 dv_2 \quad (15.37)$$

where the sums run over the occupied MOs. (The factor 4 comes from the four interorbital repulsions of the electrons in each MO pair.) We define the localized MOs as those orthonormal MOs that minimize (15.37). [It turns out that minimization of (15.37) implies the maximization of the total *intraorbital* electron repulsion and the minimization of the magnitude of the (interorbital) exchange energy.] This definition was originally suggested by Lennard-Jones and Pople and has been applied to many molecules by Edmiston and Ruedenberg [C. Edmiston and K. Ruedenberg, *Rev. Mod. Phys.*, **35**, 457 (1963); *J. Chem. Phys.*, **43**, S97 (1965); P.-O. Löwdin, ed., *Quantum Theory of Atoms, Molecules, and the Solid State*, Academic Press, 1966, pages 263–280].

The MOs that minimize (15.37) are called the **energy-localized MOs**. For molecules with symmetry, we expect that the energy-localized MOs will also be equivalent orbitals, and this is borne out by the calculated energy-localized MOs. Energy-localized MOs are thus a generalization of equivalent MOs.

Liang and Taylor calculated the energy-localized MOs for H<sub>2</sub>O, starting with the minimal-basis canonical MOs (15.19). In terms of the canonical (delocalized) MOs, Liang and Taylor found (J. H. Liang, Ph.D. thesis, Ohio State University, 1970; quoted in F. Franks, ed., *Water*, Vol. 1, Plenum, 1972, p. 42)

$$\begin{aligned} i(\text{O}) &= 0.99(1a_1) - 0.12(2a_1) + 0.06(3a_1) \\ b(\text{OH}_1) &= 0.05(1a_1) + 0.57(2a_1) + 0.71(1b_2) + 0.42(3a_1) \\ b(\text{OH}_2) &= 0.05(1a_1) + 0.57(2a_1) - 0.71(1b_2) + 0.42(3a_1) \\ l_1(\text{O}) &= 0.08(1a_1) + 0.42(2a_1) - 0.57(3a_1) - 0.71(1b_1) \\ l_2(\text{O}) &= 0.08(1a_1) + 0.42(2a_1) - 0.57(3a_1) + 0.71(1b_1) \end{aligned} \quad (15.38)$$

The lone-pair  $1b_1$  MO ( $2p_{xO}$ ) is equally divided between the equivalent lone-pair localized orbitals; note that  $0.71 = (0.5)^{1/2}$ . The inner-shell  $1a_1$  MO contributes substantially to only the  $i(\text{O})$  MO. The bonding  $1b_2$  MO is equally divided between the two bonding localized MOs. The bonding  $2a_1$  MO makes substantial contributions to the two bonding localized MOs and smaller, but still substantial, contributions to the lone-pair localized MOs. The largely lone-pair  $3a_1$  MO makes substantial contributions to the lone-pair localized MOs and lesser contributions to the bonding localized MOs.

In terms of the AOs, the energy-localized MOs of H<sub>2</sub>O are

$$\begin{aligned} i(\text{O}) &= -0.007(\text{H}_1 1s) - 0.007(\text{H}_2 1s) + 0.99(\text{O} 1s) - 0.12(\text{O} 2s_{\perp}) + 0.03(\text{O} 2p_z) \\ b(\text{OH}_1) &= 0.50(\text{H}_1 1s) - 0.10(\text{H}_2 1s) + 0.02(\text{O} 1s) + 0.25(\text{O} 2s_{\perp}) \\ &\quad + 0.407(\text{O} 2p_z) + 0.441(\text{O} 2p_y) \\ b(\text{OH}_2) &= -0.10(\text{H}_1 1s) + 0.50(\text{H}_2 1s) + 0.02(\text{O} 1s) + 0.25(\text{O} 2s_{\perp}) \\ &\quad + 0.407(\text{O} 2p_z) - 0.441(\text{O} 2p_y) \\ l_1(\text{O}) &= -0.09(\text{H}_1 1s) - 0.09(\text{H}_2 1s) + 0.09(\text{O} 1s) + 0.63(\text{O} 2s_{\perp}) \\ &\quad - 0.39(\text{O} 2p_z) - 0.71(\text{O} 2p_x) \\ l_2(\text{O}) &= -0.09(\text{H}_1 1s) - 0.09(\text{H}_2 1s) + 0.09(\text{O} 1s) + 0.63(\text{O} 2s_{\perp}) \\ &\quad - 0.39(\text{O} 2p_z) + 0.71(\text{O} 2p_x) \end{aligned} \quad (15.39)$$

## The MO wave function

$$|i(\text{O})\overline{i(\text{O})}b(\text{OH}_1)\overline{b(\text{OH}_1)}b(\text{OH}_2)\overline{b(\text{OH}_2)}l_1(\text{O})\overline{l_1(\text{O})}l_2(\text{O})\overline{l_2(\text{O})}|$$

is identical to (15.21).

Let us analyze these localized MOs for water. The  $i(\text{O})$  MO is nearly a pure  $1s$  inner-shell oxygen AO.

To define the angle between the two localized bonding MOs, we draw the line from O to  $\text{H}_1$  along which the electron probability density in  $b(\text{OH}_1)$  is a maximum, and we draw a similar line from O to  $\text{H}_2$ . The angle between these lines where they intersect at the O nucleus defines the angle between the localized bonding MOs. (If this angle differs significantly from the angle defined by straight lines between the nuclei, the bonds are said to be *bent*.) The angle between the localized bonding MOs is determined mainly by the oxygen  $2p_y$  and  $2p_z$  AO contributions (with a small influence by the hydrogen  $1s$  AOs). For  $b(\text{OH}_1)$  the  $\text{O}2p_y\text{O}2p_z$  contribution contains the terms  $0.407z_{\text{O}} + 0.441y_{\text{O}}$ . Let us rotate the coordinates in the  $zy$  plane by an angle  $\alpha = \arctan(0.441/0.407) = 47\frac{1}{2}^\circ$ , as in Fig. 15.6. The relation between coordinates in the unrotated system and in the rotated  $z'y'$  system is given by the well-known formulas

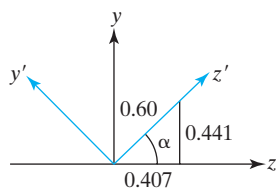


FIGURE 15.6 A coordinate rotation in the  $yz$  plane.

$$\begin{aligned} z' &= z \cos \alpha + y \sin \alpha \\ y' &= -z \sin \alpha + y \cos \alpha \end{aligned} \quad (15.40)$$

From (15.40) and Fig. 15.6 we have  $0.407z_{\text{O}} + 0.441y_{\text{O}} = 0.600z'_{\text{O}}$ . Multiplication by the exponential factor of the  $2p$  oxygen AO then gives

$$0.407(\text{O}2p_z) + 0.441(\text{O}2p_y) = 0.60(\text{O}2p_{z'}) \quad (15.41)$$

In other words, the hybridized  $2p_z2p_y$  AO on the left of (15.41) is the same function as 0.60 times a  $2p$  AO inclined at an angle of  $47\frac{1}{2}^\circ$  with the  $z$  axis. Thus the bonding  $2p_z2p_y$  hybrid AOs of oxygen in  $b(\text{OH}_1)$  and  $b(\text{OH}_2)$  point in the general direction of the hydrogen atoms. The contribution from the hydrogen atoms to  $b(\text{OH}_1)$  is mostly from  $\text{H}_11s$ , and the overlap between  $\text{H}_11s$  and the  $2p_z2p_y$  oxygen hybrid then forms the  $\text{O}-\text{H}_1$  chemical bond. The angle between the two hybrid oxygen AOs in  $b(\text{OH}_1)$  and  $b(\text{OH}_2)$  is  $95^\circ$ , and this is approximately the angle between the localized bonding MOs.

Energy-localized  $\text{H}_2\text{O}$  bonding MOs calculated from an extended-basis-set SCF MO wave function were found to have an angle of  $96^\circ$  between the oxygen hybrids that contribute to these MOs and an angle of  $103^\circ$  between the localized MOs themselves, which is nearly the same as the  $104\frac{1}{2}^\circ$  molecular bond angle, indicating negligible bending for the bonds in  $\text{H}_2\text{O}$ . [W. von Niessen, *Theor. Chim. Acta*, **29**, 29 (1973).] In contrast, in the strained molecule cyclopropane, the angle between the carbon hybrids contributing to the carbon-carbon bonding energy-localized MOs deviates outward by  $28^\circ$  from the internuclear lines [M. D. Newton, E. Switkes, and W. N. Lipscomb, *J. Chem. Phys.*, **53**, 2645 (1970)].

The  $b(\text{OH}_1)$  energy-localized MO is not completely confined to the region of the  $\text{O}-\text{H}_1$  bond. We see from (15.39) that this MO has a small contribution from the  $\text{H}_21s$  AO. (The ratio of the contributions of two AOs to an MO is given essentially by the square of the ratio of their coefficients.) Consider the contributions of  $\text{O}2s_\perp$ ,  $\text{O}2p_y$ , and  $\text{O}2p_z$  to  $b(\text{OH}_1)$ . The  $2p_y2p_z$  hybrid has a positive lobe in the region of the  $\text{OH}_1$  bond, and the contribution of the  $\text{O}2s_\perp$  AO reinforces this positive lobe. Overlap with  $\text{H}_11s$  then gives the  $\text{O}-\text{H}_1$  bond. The  $2p_y2p_z$  hybrid has a negative lobe on the side of the oxygen opposite the  $\text{OH}_1$  bond. This negative lobe is partly, but not completely, canceled by the  $\text{O}2s_\perp$  contribution. Thus the  $b(\text{OH}_1)$  MO has a “tail” on the side of the O away from  $\text{H}_1$ , this tail being somewhat distorted toward  $\text{H}_2$  by the  $-0.10\text{H}_21s$  contribution

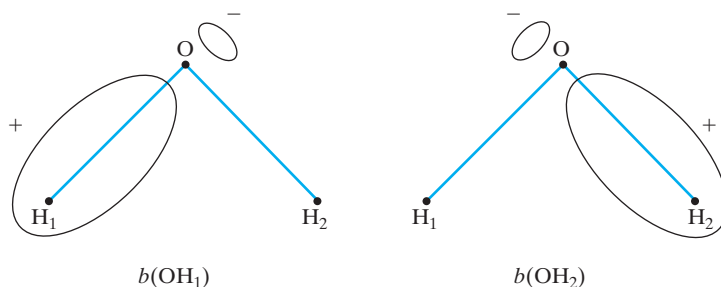


**FIGURE 15.7** Rough sketches of the localized bonding MOs in  $\text{H}_2\text{O}$ .

(Fig. 15.7). (For accurately plotted contours, see F. Franks, ed., *Water*, Vol. 1, Plenum, New York, 1972, p. 46.) Despite this tail, this MO is far more localized than any of the bonding canonical MOs in (15.19). Because  $b(\text{OH}_1)$  is mostly localized in the  $\text{OH}_1$  region, we expect only a small change in its form on going from  $\text{HOH}$  to, say,  $\text{HOCl}$ . The observed near invariance of the  $\text{O}-\text{H}$  bond from molecule to molecule is explained by MO theory using localized bonding MOs. (Neglect of the contribution of  $\text{H}_2 1s$  to the  $b(\text{OH}_1)$  localized MO gives a two-center orbital called a *bond orbital*.)

Finally, consider the two lone-pair MOs  $l_1(\text{O})$  and  $l_2(\text{O})$ . These MOs are mainly localized on the oxygen atom and are equivalent to each other. They are directed away from the hydrogen atoms and project above and below the molecular ( $yz$ ) plane (Fig. 15.8). The  $\hat{\sigma}_v(yz)$  operation interchanges the lone-pair MOs. The angle between them is approximately  $2 \arctan(0.71/0.39) = 122^\circ$ . The angle between localized lone-pair MOs calculated from a more accurate wave function is  $114^\circ$ ; W. von Niessen, *Theor. Chim. Acta*, **29**, 29 (1973).

For  $\text{N}_2$  we expect ( $:\text{N}\equiv\text{N}:$ ) the localized MOs to be an inner-shell  $1s$  AO on each atom, a lone-pair  $2s$  AO on each atom, and three bonding MOs spread over the two atoms. The canonical-MO picture is that the triple bond consists of one  $\sigma$  bond and two  $\pi$  bonds, as in Section 13.7. The energy-localized bonding MOs were found by Edmiston and Ruedenberg to be three equivalent banana-shaped orbitals spaced  $120^\circ$  apart from one another. The AOs that contribute significantly to the localized bonding MOs are the  $2s$ ,  $2p\sigma$ ,  $2p\pi_x$ , and  $2p\pi_y$  orbitals of each atom. The  $i(\text{N}_a)$  and  $i(\text{N}_b)$  localized MOs were found to be nearly pure  $1s$  nitrogen AOs. Each of the  $l(\text{N}_a)$  and  $l(\text{N}_b)$  localized MOs is a hybrid of the  $2s$  and  $2p\sigma$  AOs of the relevant nitrogen atom, with the  $2s$  AO making the larger contribution to the MO. Each lone-pair localized MO is directed away from the other nitrogen atom.

The concept of localized MOs is not as widely applicable as that of delocalized canonical MOs. Delocalized MOs are valid for any molecule. However (as noted in Section 11.5),

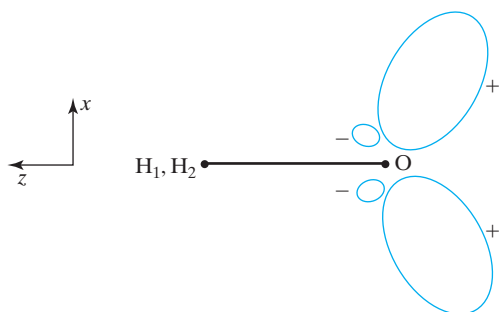


**FIGURE 15.8** Rough sketches of the localized lone-pair MOs of  $\text{H}_2\text{O}$ .

the Hartree–Fock wave functions of nonclosed-shell electronic states are, in most cases, linear combinations of a few Slater determinants [for example, see (10.44) and (10.45)], and the above localization procedure does not apply to the open-shell orbitals in these wave functions. Thus, in a molecule in an excited electronic state with an open-shell configuration, the electrons in the incompletely filled MOs are delocalized over much of the molecule.

The great success of the concept of chemical bonds between pairs of atoms is a reflection of the fact that the electronic ground states of most molecules have closed-shell configurations, for which the localized MO description is just as valid as the delocalized MO description.

For a closed-shell ground-state molecule, those properties that involve only the ground-state wave function can be calculated just as well with either the localized or the delocalized MO description. Such properties include electron probability density, dipole moment, geometry, and heat of formation. Properties of a molecule that involve the wave function of the ground state and also the wave function of an open-shell excited state or the wave function of an open-shell ion cannot be calculated using a localized MO description. Such properties include the electronic absorption spectrum and molecular ionization energies.

Localized MOs are approximately transferable from molecule to molecule, which is not true for canonical MOs. The localized  $b(\text{CH})$  MO in  $\text{CH}_4$  is quite similar to the localized  $b(\text{CH})$  MOs in  $\text{C}_2\text{H}_6$  and in  $\text{CH}_3\text{OH}$  [S. Rothenberg, *J. Chem. Phys.*, **51**, 3389 (1969); *J. Am. Chem. Soc.*, **93**, 68 (1971)].

Calculation of the Edmiston–Ruedenberg energy-localized MOs is very time consuming. Boys (and Foster) proposed a method to find localized MOs that is computationally much faster than the Edmiston–Ruedenberg method and that gives similar results in most cases; see D. A. Kleier, *J. Chem. Phys.*, **61**, 3905 (1974). The Boys method defines the LMOs as those that maximize the sum of the squares of the distances between the centroids of charge of all pairs of occupied LMOs. The *centroid of charge* of orbital  $\phi_i$  is defined as the point at  $(x_C, y_C, z_C)$ , where  $x_C \equiv \langle \phi_i | x | \phi_i \rangle$ ,  $y_C \equiv \langle \phi_i | y | \phi_i \rangle$ ,  $z_C \equiv \langle \phi_i | z | \phi_i \rangle$ . If  $r_{ij}$  is the distance between the centroids of LMOs  $i$  and  $j$ , the Boys LMOs maximize  $\sum_{j>i} \sum_i r_{ij}^2$ .

Still another way to calculate LMOs is the *Pipek–Mezey method*, which maximizes a certain sum that is related to the Mulliken gross populations of the orbitals [J. Pipek and P. G. Mezey, *J. Chem. Phys.*, **90**, 4916 (1989)]. Unlike the Edmiston–Ruedenberg and the Boys LMOs, the Pipek–Mezey LMOs for a double bond consist of one  $\sigma$  and one  $\pi$  MO. (See Section 15.9.)

As well as providing insight into chemical bonds, LMOs are useful in speeding up correlation calculations on large molecules (see Section 16.3).

## 15.9 The SCF MO Treatment of Methane, Ethane, and Ethylene

### Methane

The AOs for a minimal-basis-set SCF MO calculation of  $\text{CH}_4$  are the carbon  $1s$ ,  $2s$ ,  $2p_x$ ,  $2p_y$ , and  $2p_z$  AOs and a  $1s$  AO on each hydrogen atom. The point group of  $\text{CH}_4$  is  $\mathcal{T}_d$ . Group theory (see *Cotton* or *Schonland*) gives the possible symmetry species as  $A_1$ ,  $A_2$ ,  $E$ ,  $T_1$ , and  $T_2$ . We shall not worry about specifying the symmetry behavior that corresponds to each symmetry species and shall use the species mainly as labels for the MOs. As usual, we set up the coordinate system with the  $z$  axis coinciding with the highest-order  $C_n$  or  $S_n$  axis. For

methane this is an  $S_4$  axis (Fig. 12.5). The coordinates of the hydrogen atoms  $H_1, H_2, H_3, H_4$  are  $(q, q, q)$ ,  $(q, -q, -q)$ ,  $(-q, q, -q)$ , and  $(-q, -q, q)$ , respectively, where  $2q$  is the edge of the cube in which the molecule is inscribed (Fig. 15.9). Note the equivalence of the  $x$ ,  $y$ , and  $z$  axes.

The carbon atom is at the center of the molecule, and the carbon  $1s$  and  $2s$  AOs are each sent into themselves by every symmetry operation. These AOs transform according to the totally symmetric species  $A_1$ . The carbon  $2p_x, 2p_y$ , and  $2p_z$  AOs are given by  $x$ ,  $y$ , or  $z$  times a radial function. Their symmetry behavior is the same as that of the functions  $x$ ,  $y$ , and  $z$ , respectively. From the formulas for rotation of coordinates [Eq. (15.52)], we see that any proper rotation sends each of the functions  $x$ ,  $y$ , and  $z$  into some linear combination of  $x$ ,  $y$ , and  $z$ . Any improper rotation is the product of some proper rotation and an inversion (Prob. 12.15); the inversion simply converts each coordinate to its negative. Hence the three-carbon  $2p$  orbitals are sent into linear combinations of one another by each symmetry operation. They must therefore transform according to one of the triply degenerate symmetry species. Further investigation (which is omitted) shows the symmetry species of the  $2p$  AOs to be  $T_2$ .

Just as in  $H_2O$ , each  $1s$  hydrogen AO in  $CH_4$  does not transform according to any of the molecular symmetry species, and it is convenient to form symmetry-adapted basis functions by taking linear combinations of the  $1s$  AOs. One obvious symmetry function is

$$\chi_1 = H_1 1s + H_2 1s + H_3 1s + H_4 1s \quad (15.42)$$

Since each methane symmetry operator permutes the hydrogen  $1s$  orbitals among themselves, (15.42) is sent into itself by each symmetry operation and belongs to the totally symmetric species  $A_1$ . We need three more symmetry functions. The construction of these is not obvious without the use of group theory, and we shall simply write down the results. The remaining three orthogonal (unnormalized) symmetry-adapted basis functions can be taken as

$$\chi_2 = H_1 1s + H_2 1s - H_3 1s - H_4 1s \quad (15.43)$$

$$\chi_3 = H_1 1s - H_2 1s + H_3 1s - H_4 1s \quad (15.44)$$

$$\chi_4 = H_1 1s - H_2 1s - H_3 1s + H_4 1s \quad (15.45)$$

Each of these three functions is transformed into some linear combination of the three functions by each symmetry operation. For example, a  $\hat{C}_3$  rotation about the  $CH_1$  bond permutes the hydrogen atoms as follows:  $1 \rightarrow 1, 2 \rightarrow 3, 3 \rightarrow 4, 4 \rightarrow 2$ . The corresponding  $\hat{O}_{C_3}$  operator transforms  $\chi_2, \chi_3$ , and  $\chi_4$  into  $\chi_3, \chi_4$ , and  $\chi_2$ , respectively. These three symmetry functions therefore transform according to one of the triply degenerate symmetry species. The function  $\chi_2$  has positive signs for the hydrogen AOs with a positive  $x$  coordinate and negative signs for the hydrogen AOs with a negative  $x$  coordinate. We thus expect  $\chi_2$  to have the same symmetry behavior as the function  $x$ . Similarly,  $\chi_3$  and  $\chi_4$  behave as  $y$  and  $z$ , respectively. (As an example, note that a  $120^\circ$  counterclockwise rotation about the  $CH_1$  bond has the following effect on the three unit vectors:  $\mathbf{i} \rightarrow \mathbf{j}, \mathbf{j} \rightarrow \mathbf{k}, \mathbf{k} \rightarrow \mathbf{i}$ ; this is the same behavior shown by the functions  $\chi_2, \chi_3$ , and  $\chi_4$  for this rotation.) These three symmetry orbitals thus transform according to the same symmetry species as  $C2p_x, C2p_y$ , and  $C2p_z$ , that is, the species  $T_2$ .

The symmetry-adapted basis functions are

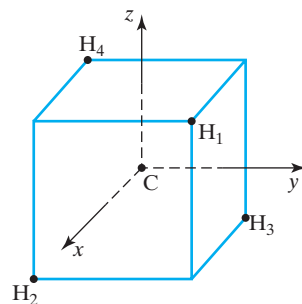


FIGURE 15.9 Coordinate axes for  $CH_4$ . The origin is at the center of the cube.

| Symmetry function | $\chi_1$ | $\chi_2$ | $\chi_3$ | $\chi_4$ | $C1s$ | $C2s$ | $C2p_x$ | $C2p_y$ | $C2p_z$ |
|-------------------|----------|----------|----------|----------|-------|-------|---------|---------|---------|
| Symmetry species  | $a_1$    | $t_2$    | $t_2$    | $t_2$    | $a_1$ | $a_1$ | $t_2$   | $t_2$   | $t_2$   |

The nine lowest MOs therefore consist of three  $a_1$  and six  $t_2$  MOs. The six  $t_2$  MOs belong to triply degenerate levels and thus fall into two different shells  $1t_2$  and  $2t_2$ . Each such shell contains three MOs of equal orbital energy and each shell holds six electrons. SCF MO calculations give the three lowest shells as  $1a_1$ ,  $2a_1$ , and  $1t_2$ , with orbital energies  $-11.20$ ,  $-0.93$ , and  $-0.54$  hartrees, respectively. The ground state of methane thus has the closed-shell electron configuration  $(1a_1)^2(2a_1)^2(1t_2)^6$  and is a  $^1A_1$  state.

Pitzer did SCF MO calculations on  $\text{CH}_4$ , using a minimal-basis set of STOs [R. M. Pitzer, *J. Chem. Phys.*, **46**, 4871 (1967)]. He found the optimized orbital exponents to be 1.17 for  $\text{H}1s$ , 5.68 for  $\text{C}1s$ , 1.76 for  $\text{C}2s$ , and 1.76 for  $\text{C}2p_z$ , which may be compared with the values 1.0, 5.7, 1.625, and 1.625 given by Slater's rules (Prob. 15.62). Pitzer found the minimum in energy to be at a bond distance of  $1.089 \text{ \AA}$ , close to the experimental value  $1.085 \text{ \AA}$ . The MOs at the experimental equilibrium bond length are

$$\begin{aligned} 1a_1 &= -0.005(\text{H}_11s + \text{H}_21s + \text{H}_31s + \text{H}_41s) + 1.001(\text{C}1s) + 0.025(\text{C}2s_{\perp}) \\ 2a_1 &= 0.186(\text{H}_11s + \text{H}_21s + \text{H}_31s + \text{H}_41s) - 0.064(\text{C}1s) + 0.584(\text{C}2s_{\perp}) \\ 1t_{2x} &= 0.318(\text{H}_11s + \text{H}_21s - \text{H}_31s - \text{H}_41s) + 0.554(\text{C}2p_x) \\ 1t_{2y} &= 0.318(\text{H}_11s - \text{H}_21s + \text{H}_31s - \text{H}_41s) + 0.554(\text{C}2p_y) \\ 1t_{2z} &= 0.318(\text{H}_11s - \text{H}_21s - \text{H}_31s + \text{H}_41s) + 0.554(\text{C}2p_z) \end{aligned} \quad (15.46)$$

The  $1a_1$  MO is essentially the carbon  $1s$  AO. The  $2a_1$  MO is a bonding combination of the carbon  $2s$  AO and the symmetry orbital (15.42). This MO has charge buildup between the carbon atom and each of the four hydrogen atoms. The  $1t_{2x}$  MO is a bonding MO; the function (15.43) is positive on the positive half of the  $x$  axis and negative on the negative half of the  $x$  axis, and its overlap with  $\text{C}2p_x$  gives charge buildup about the positive and negative sides of the  $x$  axis. Similarly, the bonding  $1t_{2y}$  and  $1t_{2z}$  MOs have charge buildup in the regions about the  $y$  and  $z$  axes, respectively.

We now consider the four localized (equivalent) bonding MOs of methane. Because of the tetrahedral symmetry, each of these orbitals must point along a  $\text{CH}$  bond, since otherwise they would not be equivalent to one another. (This is not true for water, where the equivalence requirement is satisfied by any two bonding MOs that make the same angle with the  $\text{C}_2$  axis.) Each localized bonding MO is some linear combination of the five canonical occupied MOs:

$$b(\text{CH}_1) = a(1a_1) + b(2a_1) + d(1t_{2x}) + e(1t_{2y}) + f(1t_{2z}) \quad (15.47)$$

with similar expressions for  $b(\text{CH}_2)$ ,  $b(\text{CH}_3)$ , and  $b(\text{CH}_4)$ . Since  $1a_1$  is a nonbonding low-energy inner-shell orbital, we expect  $|a| \ll |b|$ . The  $1a_1$  and  $2a_1$  canonical MOs are directed equally to all four hydrogens, and so varying  $a$  or  $b$  in (15.47) does not direct  $b(\text{CH}_1)$  preferentially to any one hydrogen. The  $1t_{2x}$ ,  $1t_{2y}$ , and  $1t_{2z}$  MOs are directed along the  $x$ ,  $y$ , and  $z$  axes, respectively, so that by adjusting  $d$ ,  $e$ , and  $f$  appropriately, we can get  $b(\text{CH}_1)$  to be localized mainly in the region of the  $\text{CH}_1$  bond. To fix  $d$ ,  $e$ , and  $f$ , we use some properties of direction cosines.

If line  $L$  passes through the origin and makes the angles  $\alpha$ ,  $\beta$ , and  $\gamma$  with the  $x$ ,  $y$ , and  $z$  axes, respectively, then the quantities

$$l \equiv \cos \alpha, \quad m \equiv \cos \beta, \quad n \equiv \cos \gamma \quad (15.48)$$

are the *direction cosines* of  $L$ . If  $(x_L, y_L, z_L)$  is a point on  $L$ , then clearly

$$x_L = r \cos \alpha, \quad y_L = r \cos \beta, \quad z_L = r \cos \gamma \quad (15.49)$$

where  $r$  is the distance from the origin. From  $x^2 + y^2 + z^2 = r^2$  it follows that

$$l^2 + m^2 + n^2 = 1 \quad (15.50)$$

Let the lines  $L_1$  and  $L_2$  go from the origin to  $(x_1, y_1, z_1)$  and  $(x_2, y_2, z_2)$ , respectively. If  $\theta_{12}$  is the angle between  $L_1$  and  $L_2$ , then [Eq. (5.20)]

$$\begin{aligned} \cos \theta_{12} &= \frac{x_1 x_2 + y_1 y_2 + z_1 z_2}{r_1 r_2} \\ \cos \theta_{12} &= l_1 l_2 + m_1 m_2 + n_1 n_2 \end{aligned} \quad (15.51)$$

Direction cosines are useful in discussing changes in coordinate axes. Let the  $x'y'z'$  and the  $xyz$  Cartesian coordinate systems have a common origin, and let the  $x'y'z'$  axes be obtained from the  $xyz$  axes by rotation, reflection, inversion, or some combination of these operations. Let the direction cosines of the  $x'$  axis with respect to the  $xyz$  system be  $l_1, m_1, n_1$ ; let  $l_2, m_2, n_2$  and  $l_3, m_3, n_3$  be the direction cosines of the  $y'$  and  $z'$  axes, respectively. Let the vector  $\mathbf{s}$  have coordinates  $(x, y, z)$  and  $(x', y', z')$  in the unrotated and rotated coordinate systems. We have  $x' = \mathbf{s} \cdot \mathbf{i}'$ , where  $\mathbf{i}'$  is a unit vector along the  $x'$  axis. Since  $\mathbf{i}'$  is of unit length, it follows from (15.49) and (15.48) that the coordinates of  $\mathbf{i}'$  in the  $xyz$  system are  $l_1, m_1$ , and  $n_1$ . Hence

$$\begin{aligned} x' &= l_1 x + m_1 y + n_1 z \\ y' &= l_2 x + m_2 y + n_2 z \\ z' &= l_3 x + m_3 y + n_3 z \end{aligned} \quad (15.52)$$

where the  $y'$  and  $z'$  equations follow from  $y' = \mathbf{s} \cdot \mathbf{j}'$  and  $z' = \mathbf{s} \cdot \mathbf{k}'$ . [Equation (15.40) is a special case of (15.52).] Since the angle between any pair of the  $x'$ ,  $y'$ , and  $z'$  axes is  $90^\circ$ , it follows from (15.51) that the direction cosines of these axes satisfy

$$\begin{aligned} l_1 l_2 + m_1 m_2 + n_1 n_2 &= 0 \\ l_1 l_3 + m_1 m_3 + n_1 n_3 &= 0 \\ l_2 l_3 + m_2 m_3 + n_2 n_3 &= 0 \end{aligned} \quad (15.53)$$

Now we return to the determination of  $d$ ,  $e$ , and  $f$ . The MOs  $1t_{2x}$ ,  $1t_{2y}$ , and  $1t_{2z}$  are directed along the  $x$ ,  $y$ , and  $z$  axes, respectively, and the contributions of the carbon  $2p_x$ ,  $2p_y$ , and  $2p_z$  AOs to these MOs are

$$xe^{-\zeta r}, \quad ye^{-\zeta r}, \quad ze^{-\zeta r} \quad (15.54)$$

The contribution of the hydrogen AOs to the  $1t_2$  MOs has a more complicated form (the hydrogens are not at the coordinate origin), but we need not explicitly consider the hydrogen part of the MOs. This is because the hydrogen symmetry orbitals (15.43) to (15.45) have the same directional properties as the corresponding carbon  $2p$  AOs (15.54) with which each is combined in the  $1t_2$  MOs [Eq. (15.46)]. The linear combination (15.47) has as its carbon  $2p$  contribution

$$(dx + ey + fz)e^{-\zeta r} \quad (15.55)$$

Let  $l_1, m_1$ , and  $n_1$  be the direction cosines of the  $\text{CH}_1$  line. We assert that if  $d$ ,  $e$ , and  $f$  are chosen as proportional to these direction cosines, then  $b(\text{CH}_1)$  will be directed toward  $\text{H}_1$ . To verify this, we set  $d : e : f = l_1 : m_1 : n_1$  in (15.55) and use (15.52) to get

$$c(l_1 x + m_1 y + n_1 z)e^{-\zeta r} = cx'e^{-\zeta r} \quad (15.56)$$

where  $c$  is some constant and the  $x'$  axis runs from C to  $\text{H}_1$ .

Similarly, by picking  $d$ ,  $e$ , and  $f$  proportional to the direction cosines of the other CH lines, we form localized orbitals along these bonds. From (15.49) the direction cosines of the CH lines are

$$\begin{aligned} \text{CH}_1: & 3^{-1/2}, 3^{-1/2}, 3^{-1/2} & \text{CH}_2: & 3^{-1/2}, -3^{-1/2}, -3^{-1/2} \\ \text{CH}_3: & -3^{-1/2}, 3^{-1/2}, -3^{-1/2} & \text{CH}_4: & -3^{-1/2}, -3^{-1/2}, 3^{-1/2} \end{aligned} \quad (15.57)$$

To satisfy the equivalence requirement, the values of  $a$  and  $b$  in (15.47) must be the same for each bonding localized MO. The equivalent localized MOs for methane thus have the forms

$$\begin{aligned} b(\text{CH}_1) &= a(1a_1) + b(2a_1) + 3^{-1/2}c(1t_{2x} + 1t_{2y} + 1t_{2z}) \\ b(\text{CH}_2) &= a(1a_1) + b(2a_1) + 3^{-1/2}c(1t_{2x} - 1t_{2y} - 1t_{2z}) \\ b(\text{CH}_3) &= a(1a_1) + b(2a_1) + 3^{-1/2}c(-1t_{2x} + 1t_{2y} - 1t_{2z}) \\ b(\text{CH}_4) &= a(1a_1) + b(2a_1) + 3^{-1/2}c(-1t_{2x} - 1t_{2y} + 1t_{2z}) \end{aligned} \quad (15.58)$$

Orthonormality of the bonding localized MOs (15.58) requires that

$$a^2 + b^2 + c^2 = 1 \quad \text{and} \quad a^2 + b^2 - \frac{1}{3}c^2 = 0 \quad (15.59)$$

Hence

$$c = \frac{1}{2}\sqrt{3}, \quad (a^2 + b^2)^{1/2} = \frac{1}{2} \quad (15.60)$$

The equivalence, direction, and orthonormality requirements have fixed all but one parameter (the ratio  $a/b$ ) in the localized bonding MOs of methane.

The  $1t_{2x}$  MO points equally in the  $+x$  and  $-x$  directions. Similarly, the  $1t_{2y}$  and  $1t_{2z}$  MOs point equally on both sides of the carbon atom. This is not true of the bonding localized MOs: The  $2a_1$  MO is positive in most of the bonding region between the carbon atom and the hydrogen atoms. [It is negative in the region very near the carbon atom, because of the  $-0.06(\text{C}1s)$  term and the negative portion of the orthogonalized  $\text{C}2s$  AO; see Eq. (15.20).] The linear combination  $1t_{2x} + 1t_{2y} + 1t_{2z}$  points equally in the  $(1, 1, 1)/\sqrt{3}$  and the  $(-1, -1, -1)/\sqrt{3}$  directions. If  $b$  in (15.58) is taken as positive (as we have taken  $c$ ), then the  $2a_1$  MO adds to the positive half of this linear combination of the  $1t_2$  MOs and cancels much of the negative half of this linear combination. With  $b$  and  $c$  having the same sign, the  $b(\text{CH}_1)$  MO points mostly in the  $(1, 1, 1)/\sqrt{3}$  direction, with only a small “tail” in the  $(-1, -1, -1)/\sqrt{3}$  direction.

Pitzer’s SCF calculation gives the energy-localized bonding and inner-shell methane MOs as

$$\begin{aligned} b(\text{CH}_1) &= 0.055(1a_1) + 0.497(2a_1) + \frac{1}{2}(1t_{2x} + 1t_{2y} + 1t_{2z}) \\ b(\text{CH}_2) &= 0.055(1a_1) + 0.497(2a_1) + \frac{1}{2}(1t_{2x} - 1t_{2y} - 1t_{2z}) \\ b(\text{CH}_3) &= 0.055(1a_1) + 0.497(2a_1) + \frac{1}{2}(-1t_{2x} + 1t_{2y} - 1t_{2z}) \\ b(\text{CH}_4) &= 0.055(1a_1) + 0.497(2a_1) + \frac{1}{2}(-1t_{2x} - 1t_{2y} + 1t_{2z}) \\ i(\text{C}) &= 0.994(1a_1) - 0.111(2a_1) \end{aligned} \quad (15.61)$$

From (15.46) and (15.61), we have

$$\begin{aligned} i(\text{C}) &= 1.002(\text{C}1s) - 0.040(\text{C}2s_{\perp}) - 0.025(\text{H}_11s + \text{H}_21s + \text{H}_31s + \text{H}_41s) \\ b(\text{CH}_1) &= 0.024(\text{C}1s) + 0.292(\text{C}2s_{\perp}) + 0.569(\text{H}_11s) - 0.066(\text{H}_21s + \text{H}_31s + \text{H}_41s) \\ &\quad + 0.277(\text{C}2p_x + \text{C}2p_y + \text{C}2p_z) \end{aligned} \quad (15.62)$$

From Eqs. (15.54) to (15.56), the linear combination  $dp_x + ep_y + fp_z$  is equal to a rotated  $p$  orbital pointing along the  $x'$  axis and containing the factor  $dx + ey + fz = c(l_1x + m_1y + n_1z) = cx'$ , where  $l_1, m_1, n_1$  are the direction cosines of the  $x'$  axis relative to the  $x, y, z$  axes. Substitution of  $l_1 = d/c, m_1 = e/c, n_1 = f/c$  into  $l_1^2 + m_1^2 + n_1^2 = 1$  [Eq. (15.50)] gives  $c = (d^2 + e^2 + f^2)^{1/2}$ . Therefore, a localized MO that contains the terms  $aC2s_\perp + dC2p_x + eC2p_y + fC2p_z$  can be viewed as containing the terms  $aC2s_\perp + (d^2 + e^2 + f^2)^{1/2}C2p_{x'}$ , and we say the hybridization of the carbon AO in this MO is

$$sp^{(d^2+e^2+f^2)/a^2} \quad (15.63)$$

For example, for the localized bonding MO (15.62) of  $\text{CH}_4$ ,  $a^2 = 0.0853$ ,  $d^2 + e^2 + f^2 = 0.230$ , and this MO is an  $sp^{2.7}$  hybrid, which is close to the traditional  $sp^3$  hybridization of VB theory (Section 16.12).

## Ethane

The most fascinating property of  $\text{C}_2\text{H}_6$  is the barrier to internal rotation about the carbon-carbon single bond. The staggered conformation of a  $\text{C}_2\text{H}_6$  molecule is more stable than the eclipsed conformation by  $0.125 \text{ eV} = 0.00461 \text{ hartree}$ , which is equivalent to  $2.89 \text{ kcal/mol}$ . In 1936, J. D. Kemp and Kenneth Pitzer discovered this fact in examining thermodynamic data for ethane. The correlation energy for an  $N$ -electron species is very roughly  $-(0.04 \text{ hartree})(N - 1)$  [S. Kristyán, *Chem. Phys. Lett.*, **247**, 101 (1995)], so the Hartree-Fock energy of  $\text{C}_2\text{H}_6$  will differ by about  $0.7 \text{ hartree} \approx 20 \text{ eV}$  from the true molecular energy. This SCF energy error is far greater than the barrier height. At first sight, it seems hopeless to expect an SCF MO calculation to give a meaningful result for the ethane barrier.

The minimal-basis-set AOs for ethane are the hydrogen  $1s$  orbitals and the carbon  $1s, 2s$ , and  $2p$  orbitals, a total of  $6(1) + 2(5) = 16$  basis AOs. To calculate the barrier in ethane, we must calculate the energy of the staggered and the eclipsed conformations, which requires two separate SCF calculations. One first forms appropriate linear combinations of the hydrogen AOs and of the carbon AOs to get symmetry-adapted basis functions. The Roothaan equations are then solved iteratively to give the basis-function coefficients and orbital energies, and the total molecular energy is found.

The pioneering ethane calculation is by Russell Pitzer (Kenneth Pitzer's son) and W. N. Lipscomb, who did an SCF MO calculation using a minimal STO basis set with orbital exponents chosen according to Slater's rules [R. M. Pitzer and W. N. Lipscomb, *J. Chem. Phys.*, **39**, 1995 (1963)]. They used the experimentally observed equilibrium geometry for staggered  $\text{C}_2\text{H}_6$  and assumed the geometry of the eclipsed form to be that given by rigidly rotating one methyl group with respect to the other. Their calculated energies of  $-78.98593$  and  $-78.99115$  hartrees for the eclipsed and staggered forms, respectively, give a barrier of  $3.3 \text{ kcal/mol}$ , in reasonable agreement with experiment. Clementi and Popkie did a near-Hartree-Fock calculation with a basis set of 102 CGTFs [E. Clementi and H. Popkie, *J. Chem. Phys.*, **57**, 4870 (1972)]. They varied the bond angles and the CC bond length for each conformation to minimize the energy and found an increase of  $0.02 \text{ Å}$  in the CC length and a decrease of  $0.3^\circ$  in the HCH angle on going from the staggered to the eclipsed form. Their calculated barrier is  $3.2 \text{ kcal/mol}$ .

Why do SCF MO calculations give good values of the ethane rotational barrier? The answer lies in the correlation energy. Electrons paired in the same localized orbital move through the same region of space. Hence the intraorbital correlation for such a pair should be substantially greater in magnitude than the interorbital correlation energy for two electrons in different localized MOs. In line with this, it was formerly believed that the magnitude of the total interorbital molecular correlation energy was much less than the total

intraorbital correlation energy. However, there are many more interorbital pair correlations than intraorbital correlations, and it is now recognized that the total interorbital correlation is not negligible and in some cases can be of comparable magnitude to the total intraorbital correlation. [See E. Steiner, *J. Chem. Phys.*, **54**, 1114 (1971).] Hence we must consider both kinds of correlation in ethane.

Rotation of a methyl group in ethane does not change any of the bonds, and thus intraorbital correlation should be essentially the same in the staggered and eclipsed forms. Moreover, we expect most of the interorbital correlations to be essentially unchanged in the two forms. In particular, correlations between the C—H bonds within each methyl group should remain virtually the same, and so should correlations between the C—C and C—H bonding pairs. It is only the correlations between C—H pairs of different methyl groups that should change, and these make the smallest contributions to interorbital correlation. Thus we expect the total correlation energy to be only slightly changed from staggered to eclipsed. Thus the energy error in an SCF MO calculation is about the same for the two forms, and the Hartree–Fock energy difference should yield a good estimate of the barrier. Recall that Hartree–Fock calculations give poor values for dissociation energies. This is because the number of electron pairs changes in forming a chemical bond from atoms, thereby changing the correlation energy substantially.

Some HF/6-31G\* results for other rotational barriers calculated with geometry optimization of all structures are (values in kcal/mol): CH<sub>3</sub>OH, 1.36 calculated versus 1.07 experimental; CH<sub>3</sub>CHO, 1.03 calculated versus 1.17 experimental; CH<sub>3</sub>NH<sub>2</sub>, 2.39 calculated versus 1.98 experimental; CH<sub>3</sub>SiH<sub>3</sub>, 1.40 calculated versus 1.70 experimental (CCCBDB).

The physical origin of the ethane rotational barrier has been the subject of intense controversy. The steric-repulsion viewpoint attributes the barrier to the Pauli repulsion (Section 10.3) between the eclipsing localized C—H bonding electron pairs in eclipsed ethane [R. M. Pitzer, *Acc. Chem. Res.*, **16**, 207 (1983); F. M. Bickelhaupt and E. J. Baerends, *Angew. Chem. Int. Ed.*, **42**, 4183 (2003)]. The hyperconjugation viewpoint [A. E. Reed and F. Weinhold, *Isr. J. Chem.*, **31**, 277 (1991); V. Pophristic and L. Goodman, *Nature*, **411**, 565 (2001); F. Weinhold, *Nature*, **411**, 539 (2001); F. Weinhold, *Angew. Chem. Int. Ed.*, **42**, 4188 (2003)] attributes the barrier to greater stabilization of the staggered form by hyperconjugation. (In the language of the VB method, hyperconjugation stabilization in ethane is due to contributions of ionic resonance structures that have a double bond between the carbons and have an H<sup>+</sup> on one carbon and an H<sup>−</sup> on the other carbon, with no bonds to these ionic H's. In the language of MO theory, hyperconjugation stabilization is due to interaction between a filled, localized bonding CH MO on one C and a vacant, localized antibonding CH MO on the second carbon. *Hyperconjugation* involves delocalization of localized  $\sigma$  electrons, whereas *conjugation* involves delocalization of localized  $\pi$  electrons.  $\sigma$  and  $\pi$  LMOs are defined at the end of Section 15.9.)

An analysis of ab initio VB wave functions for ethane and related molecules found that the main contributor to the barrier is steric repulsion, with hyperconjugation contributing only one-third of the ethane barrier [Y. Mo et al., *Angew. Chem. Int. Ed.*, **43**, 1986 (2004); L. Song et al., *J. Phys. Chem. A*, **109**, 2310 (2005); Y. Mo and J. Gao, *Acc. Chem. Res.*, **40**, 113 (2007); Y. Mo, *WIREs Comput. Mol. Sci.*, **1**, 164 (2011)]. The experimentally observed structure of CH<sub>3</sub>CH<sub>2</sub>CCCH<sub>2</sub>CH<sub>3</sub> has been interpreted as providing strong support for the steric-repulsion origin of the ethane barrier [R. K. Bohn, *J. Phys. Chem. A*, **108**, 6814 (2004)].

### Ethylene

For C<sub>2</sub>H<sub>4</sub>, the ground-state equilibrium-geometry point group is  $\mathcal{D}_{2h}$ . The standard choice of coordinate axes is shown in Fig. 15.10.

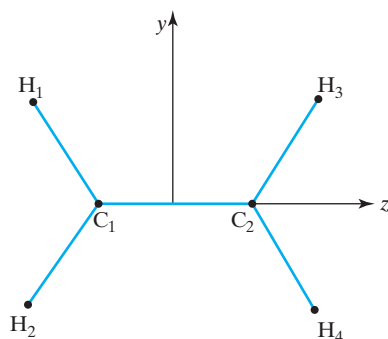


**FIGURE 15.10** Coordinate axes for C<sub>2</sub>H<sub>4</sub>. The x axis is perpendicular to the molecular plane.

There are eight symmetry operations for  $\mathcal{D}_{2h}$ . Each symmetry operation commutes with every other symmetry operation, so the electronic wave function must be an eigenfunction of all the symmetry operators, and we have only nondegenerate (*A* and *B*) symmetry species. Since the three rotations, the three reflections, and the inversion operation each have their squares equal to the identity operation, these symmetry operations must have the eigenvalues  $\pm 1$ . All eight symmetry operations can be expressed as the product of one, two, or three of the reflections (each reflection simply converts one coordinate to its negative):

$$\hat{E} = [\hat{\sigma}(xy)]^2, \quad i = \hat{\sigma}(xy)\hat{\sigma}(xz)\hat{\sigma}(yz)$$

$$\hat{C}_2(x) = \hat{\sigma}(xy)\hat{\sigma}(xz), \quad \hat{C}_2(y) = \hat{\sigma}(xy)\hat{\sigma}(yz), \quad \hat{C}_2(z) = \hat{\sigma}(xz)\hat{\sigma}(yz)$$

Since the symmetry eigenvalues multiply the same way the symmetry operations do, specification of the eigenvalues of the three reflections is sufficient to fix all eight symmetry eigenvalues. There are thus  $2^3 = 8$  possible symmetry species. The standard notation for these is given in [Table 15.3](#). The *g* or *u* subscript corresponds to eigenvalue +1 or -1 for  $\hat{i}$ . For  $\mathcal{D}_{2h}$  the designation *A* is used only for symmetry species with eigenvalue +1 for all three  $\hat{C}_2$  rotations.

There are 14 minimal-basis-set AOs. It is easy to set up symmetry orbitals by trial and error. Unnormalized symmetry orbitals and their symmetry species are listed in [Table 15.4](#).

The 14 lowest MOs include four *a<sub>g</sub>*, four *b<sub>1u</sub>*, two *b<sub>2u</sub>*, two *b<sub>3g</sub>*, one *b<sub>3u</sub>*, and one *b<sub>2g</sub>* MO. Eight of these MOs are occupied in the ground state. SCF MO calculations give as the electron configuration of the <sup>1</sup>*A<sub>g</sub>* ground state [U. Kaldor and I. Shavitt, *J. Chem. Phys.*, **48**, 191 (1968)]

$$(1a_g)^2(1b_{1u})^2(2a_g)^2(2b_{1u})^2(1b_{2u})^2(3a_g)^2(1b_{3g})^2(1b_{3u})^2$$

*Each canonical MO of a planar molecule is classified as  $\sigma$  or  $\pi$  according to whether its eigenvalue for reflection in the molecular plane is +1 or -1, respectively.* (This usage is somewhat inconsistent with the  $\sigma$ ,  $\pi$ ,  $\delta$ ,  $\cdots$  classification of linear-molecule MOs. For linear molecules, the symbols  $\sigma$  and  $\pi$  signify an axial component of electronic orbital angular momentum of 0 and  $\pm\hbar$ , respectively. For nonlinear molecules we cannot specify the component of **L** along an internuclear line. For linear molecules,  $\sigma$  MOs are nondegenerate and  $\pi$  MOs are doubly degenerate. For nonlinear molecules, the  $\sigma$ - $\pi$  classification is unrelated to the degeneracy.) For the ground state of water, all the occupied MOs are  $\sigma$  MOs except the lone-pair  $1b_1$  MO, which is a  $\pi$  MO.

**TABLE 15.3 Symmetry Species Eigenvalues for  $\mathcal{D}_{2h}$** 

|          | $\hat{\sigma}(xy)$ | $\hat{\sigma}(xz)$ | $\hat{\sigma}(yz)$ |
|----------|--------------------|--------------------|--------------------|
| $A_g$    | +1                 | +1                 | +1                 |
| $A_u$    | -1                 | -1                 | -1                 |
| $B_{1g}$ | +1                 | -1                 | -1                 |
| $B_{1u}$ | -1                 | +1                 | +1                 |
| $B_{2g}$ | -1                 | +1                 | -1                 |
| $B_{2u}$ | +1                 | -1                 | +1                 |
| $B_{3g}$ | -1                 | -1                 | +1                 |
| $B_{3u}$ | +1                 | +1                 | -1                 |

**TABLE 15.4 Symmetry-Adapted Basis Functions for  $C_2H_4$** 

| Symmetry Function               | Symmetry Species |
|---------------------------------|------------------|
| $H_11s + H_21s + H_31s + H_41s$ | $a_g$            |
| $C_11s + C_21s$                 | $a_g$            |
| $C_12s + C_22s$                 | $a_g$            |
| $C_12p_z - C_22p_z$             | $a_g$            |
| $H_11s + H_21s - H_31s - H_41s$ | $b_{1u}$         |
| $C_11s - C_21s$                 | $b_{1u}$         |
| $C_12s - C_22s$                 | $b_{1u}$         |
| $C_12p_z + C_22p_z$             | $b_{1u}$         |
| $H_11s - H_21s + H_31s - H_41s$ | $b_{2u}$         |
| $C_12p_y + C_22p_y$             | $b_{2u}$         |
| $H_11s - H_21s - H_31s + H_41s$ | $b_{3g}$         |
| $C_12p_y - C_22p_y$             | $b_{3g}$         |
| $C_12p_x + C_22p_x$             | $b_{3u}$         |
| $C_12p_x - C_22p_x$             | $b_{2g}$         |

For ethylene, the only occupied  $\pi$  MO in the ground electronic state is the  $1b_{3u}$  MO, the highest occupied MO. Since there is only one  $b_{3u}$  symmetry orbital in Table 15.4, the minimal-basis  $1b_{3u}$  MO must be identical (apart from a normalization constant) to this symmetry orbital. One finds (using STOs)  $1b_{3u} = 0.63(C_12p_x + C_22p_x)$ . This bonding MO formed by overlap of the  $2p_x$  AOs of carbon resembles the  $\pi_u$  MO of Fig. 13.15. The  $1b_{3u}$   $\pi$  MO accounts for the planarity of ethylene in its ground state. As one  $CH_2$  group is rotated relative to the other, the overlap between the two carbon  $2p_x$  AOs rapidly diminishes, and the energy of the  $1b_{3u}$  MO increases; hence the molecule strongly resists torsion about the carbon–carbon bond.

The lowest-lying unoccupied MO of ethylene is the  $1b_{2g}$  antibonding  $\pi$  MO:  $1b_{2g} = 0.82(C_12p_x - C_22p_x)$ . It resembles the  $\pi_g$  MO in Fig. 13.15. The excited ethylene configuration  $(1a_g)^2 \cdots (1b_{3g})^2(1b_{3u})(1b_{2g})$  gives rise to two terms (a singlet and a triplet), and it is likely that these electronic states are nonplanar, with one  $CH_2$  group rotated about  $90^\circ$  with respect to the other [K. B. Wiberg et al., *J. Phys. Chem.*, **96**, 10756 (1992)]. Moreover, in the singlet excited state, one of the  $CH_2$  groups is pyramidalized, with the two CH bonds of one  $CH_2$  group bent toward one of the hydrogens of the other  $CH_2$  group, so that the plane of one  $CH_2$  group is not perpendicular to the plane of the other  $CH_2$  group [S. El-Taher et al., *Int. J. Quantum Chem.*, **82**, 242 (2001); M. Barbatti et al., *J. Chem. Phys.* **121**, 11614 (2004).]

The canonical  $\sigma$  MOs of ethylene each contain contributions from AOs of all six atoms and are delocalized over the whole molecule. By taking appropriate linear combinations of the canonical  $\sigma$  MOs, we can form localized  $\sigma$  MOs. We expect these MOs to consist of the following inner-shell and bonding MOs:

$$i(C_1), \quad i(C_2) \\ b(C_1C_2), \quad b(C_1H_1), \quad b(C_1H_2), \quad b(C_2H_3), \quad b(C_2H_4) \quad (15.64)$$

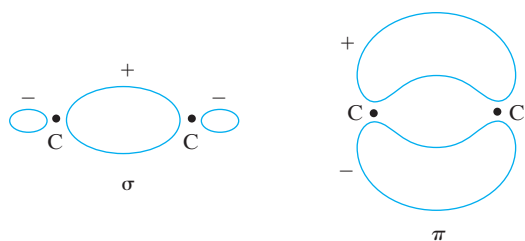


**FIGURE 15.11**  $\sigma$ ,  $\pi$  description of the ethylene double bond. The cross sections are taken in a plane perpendicular to the molecular plane, which is a nodal plane for the  $\pi$  MO.

The MOs  $b(\text{C}_1\text{C}_2)$  and  $1b_{3u}$  give the familiar description of the carbon–carbon double bond as one  $\sigma$  and one  $\pi$  bond (Fig. 15.11). We still do not, however, have the equivalent MOs of ethylene, since the  $\pi$  MO  $1b_{3u}$  is not equivalent to any of the  $\sigma$  MOs in (15.64), nor is it equivalent to itself; it goes into its negative upon a  $\hat{C}_2(z)$  rotation. By adding and subtracting the  $b(\text{C}_1\text{C}_2)$  and  $1b_{3u}$  MOs, we can form two equivalent localized carbon–carbon bonding MOs:

$$b_1(\text{C}_1\text{C}_2) = 2^{-1/2}[b(\text{C}_1\text{C}_2) + 1b_{3u}] \quad (15.65)$$

$$b_2(\text{C}_1\text{C}_2) = 2^{-1/2}[b(\text{C}_1\text{C}_2) - 1b_{3u}] \quad (15.66)$$

The MOs (15.65) and (15.66) lead to the description of the carbon–carbon bond as composed of two “banana” bonds (Fig. 15.12).

There has been controversy as to whether the ethylene double bond is best described as two equivalent bent banana bonds or as a  $\sigma$  and a  $\pi$  bond. Kaldor calculated the energy-localized MOs of ethylene from minimal-basis-set SCF MOs [U. Kaldor, *J. Chem. Phys.*, **46**, 1981 (1967)]. Since there is no a priori necessity that the energy-localized MOs be equivalent orbitals, this calculation provides evidence as to which is the “better” description of the carbon–carbon double bond. Kaldor found the energy-localized carbon–carbon bond orbitals in ethylene to be the two equivalent banana bonds. For acetylene he found the energy-localized carbon–carbon bond orbitals to be three equivalent banana bonds and not one  $\sigma$  and two  $\pi$  bonds. Of course, although the electron densities in the individual MOs differ for the banana-bond versus  $\sigma$ – $\pi$  descriptions, the total probability density for the four or six electrons in the double or triple bond is the same in either picture.

The energy-localized  $\text{C}_1\text{H}_1$  bond MO in ethylene was found by Kaldor to be

$$0.3637(\text{C}_1 2s_\perp) + 0.4143(\text{C}_1 2p_y) - 0.2574(\text{C}_1 2p_z) + 0.4939(\text{H}_1 1s) + \dots$$

where the dots indicate small contributions from other AOs. From (15.63), the hybridization of carbon in this localized MO is  $sp^{1.8}$ , which is close to the  $sp^2$  of VB theory. The angle made by the carbon hybrid AO in  $b(\text{CH}_1)$  with the C–C axis is  $\pi - \arctan(1.61) = 122^\circ$ , essentially the same as the experimental bond angle. Hence the localized CH bond orbitals are not bent in planar ethylene.

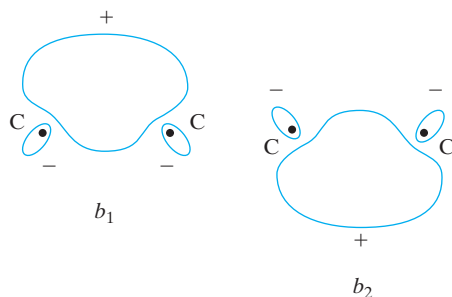


**FIGURE 15.12** Banana-bond description of the ethylene double bond. The cross sections are the same as in Fig. 15.11.

We noted earlier in this section that each *canonical* MO of a planar molecule is classified as  $\pi$  or  $\sigma$  according to whether or not the molecular plane is a nodal plane for the MO. A *localized* bonding MO of a molecule (planar or nonplanar) is classified as  $\sigma$  or  $\pi$  or  $\delta$  according to whether this MO has 0 or 1 or 2 nodal planes containing the line joining the nuclei of the two bonded atoms. The bonding localized MOs in water (Fig. 15.7) are  $\sigma$  MOs; the ethylene double-bond localized MOs in Fig. 15.11 consist of one  $\sigma$  and one  $\pi$  MO; the  $b(\text{CH}_1)$  MO (15.62) in  $\text{CH}_4$  is a  $\sigma$  MO. In certain transition-metal compounds (for example,  $\text{Re}_2\text{Cl}_8^{2-}$ ), overlap of  $d$  AOs produces a  $\delta$  localized bonding MO, and these compounds have a quadruple bond consisting of one  $\sigma$ , two  $\pi$ , and one  $\delta$  MO; see F. A. Cotton et al. (eds.), *Multiple Bonds Between Metal Atoms*, 3rd ed., Springer, 2005. A few organometallic compounds exist that contain quintuple bonds between two Cr atoms or between two Mo atoms ([en.wikipedia.org/wiki/Quintuple\\_bond](http://en.wikipedia.org/wiki/Quintuple_bond)) and the gas-phase dimers  $\text{Mo}_2$  and  $\text{W}_2$  contain sextuple bonds ([en.wikipedia.org/wiki/Sextuple\\_bond](http://en.wikipedia.org/wiki/Sextuple_bond)).

## 15.10 Molecular Geometry

### Equilibrium Geometry

The equilibrium geometry of a molecule corresponds to the nuclear arrangement that minimizes  $U$ , the molecular electronic energy including internuclear repulsion [Eqs. (13.4) and (13.8)].

The changes in  $U$  as a bond length or bond angle varies over moderate ranges are substantially smaller in magnitude than the molecular correlation energy. For example, for  $\text{H}_2\text{O}$  a large basis-set CI calculation [P. Hennig et al., *Theor. Chim. Acta*, **47**, 233 (1978)] found that a variation of  $\pm 15^\circ$  from the equilibrium  $105^\circ$  angle changed  $U$  by 0.008% and variations of  $-0.16 \text{ \AA}$  or  $+0.21 \text{ \AA}$  from the equilibrium OH bond lengths changed  $U$  by 0.07%. In contrast, the correlation energy of  $\text{H}_2\text{O}$  is 0.5% of  $U$  (Table 15.2). Also, changes in dihedral angles produce changes that are far less than the correlation energy (recall the discussion of ethane in Section 15.9).

Despite the smallness of these changes of molecular energy with nuclear locations as compared with the energy error (the correlation energy) inherent in the SCF MO method, ab initio SCF MO wave functions usually give good predictions (0 to 3% error) of equilibrium bond distances and angles in molecules not involving transition metals. Some HF/6-31G\* bond angles and lengths follow (taken mostly from the CCCBDB), where experimental values are in parentheses.  $\text{NH}_3$ :  $107^\circ$  ( $107^\circ$ ) and  $1.00 \text{ \AA}$  ( $1.01 \text{ \AA}$ );  $\text{H}_2\text{O}$   $105.5^\circ$  ( $104.5^\circ$ ) and  $0.95 \text{ \AA}$  ( $0.96 \text{ \AA}$ );  $\text{C}_2\text{H}_6$ :  $108^\circ$  ( $108^\circ$ ) for HCH,  $1.09 \text{ \AA}$  ( $1.09 \text{ \AA}$ ) for C—H, and  $1.53 \text{ \AA}$  ( $1.54 \text{ \AA}$ ) for C—C;  $\text{C}_2\text{H}_4$ :  $116^\circ$  ( $117^\circ$ ) for HCH and  $1.32 \text{ \AA}$  ( $1.34 \text{ \AA}$ ) for C=C;  $\text{H}_2\text{CO}$ :  $116^\circ$  ( $116^\circ$ ) for HCH and  $1.18 \text{ \AA}$  ( $1.20 \text{ \AA}$ ) for C=O. (For extensive comparisons, see the CCCBDB and *Hehre*.)

Evidently, the correlation energy remains approximately constant for bond angle and length variations in the region of the equilibrium geometry.

For transition-metal compounds, HF geometries are not reliable and one of the correlation methods of Chapter 16 needs to be used. Also, HF/STO-3G geometries for compounds not containing transition elements fairly often show substantial errors and should not be relied on. The minimal STO-3G basis set is essentially obsolete because of its unreliable performance for molecular properties.

### The Potential-Energy Surface (PES)

The geometry of a nonlinear molecule with  $N$  nuclei is defined by  $3N - 6$  independent nuclear coordinates  $q_1, q_2, \dots, q_{3N-6}$ , and its electronic energy  $U$  is a function of these coordinates. The number 6 is subtracted from the total number of nuclear coordinates because the three translational and three rotational degrees of freedom leave  $U$  unchanged.

(A diatomic molecule has only two rotational degrees of freedom, the angles  $\theta$  and  $\phi$  in Fig. 6.3, and here  $U$  is a function of only one variable, the internuclear distance  $R$ .) The function  $U$  gives what is called the **potential-energy surface** (PES) for the molecule, so called because  $U$  is the potential energy in the Schrödinger equation (13.10) and (13.11) for nuclear motion. If  $U$  depended on two variables, then a plot of  $U(q_1, q_2)$  in three dimensions would give a surface in ordinary three-dimensional space. Because of the large number of variables,  $U$  is a “surface” in an abstract “space” of  $3N - 5$  dimensions. To find  $U$ , we must solve the electronic Schrödinger equation at many nuclear configurations, which is a formidable task for a large molecule. Calculation of  $U$  at one particular arrangement of the nuclei is called a **single-point calculation**, since it gives one point on the molecular PES.

A complicating fact is that a large molecule may have many minima on its PES. **Figure 15.13** sketches the variation of the electronic energy  $U$  for butane,  $\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_3$ , as a function of the CCCC dihedral angle. At each point on this curve, all geometrical coordinates except the CCCC dihedral angle have been varied to yield the energy minimum for the particular fixed CCCC dihedral angle. The  $0^\circ$  dihedral angle gives the **syn** (or *cis*) conformation with the methyl groups eclipsing one another. This is an energy maximum with respect to variation of the dihedral angle. However, because the geometry has been optimized for all variables except the dihedral angle, the  $0^\circ$  point corresponds to an energy minimum for the  $3N - 7$  remaining variables. The  $0^\circ$  point is a **first-order saddle point**, meaning that it is an energy maximum for one variable and an energy minimum for the remaining variables. (The point where the rider sits on a saddle is a point of maximum height on the curve that goes from one side of the saddle to the other and is a minimum point on the curve that goes from the back to the front of the saddle.) The energy minimum at about  $65^\circ$  corresponds to the **+gauche** conformation, and the minimum at  $180^\circ$  is the **anti** (or *trans*) conformation. The  $180^\circ$  minimum is the lowest-energy point on the butane PES and so is called the **global minimum**. The  $65^\circ$  minimum is a **local minimum**, meaning that it is lower in energy than all PES points in its immediate vicinity. The minimum near  $-65^\circ$  is the **-gauche** conformation, which is a nonsuperimposable mirror image of the +gauche conformation.

The **conformation** of a molecule is specified by giving the values of all dihedral angles about single bonds. A conformation that corresponds to an energy minimum (local or global) is called a **conformer**.

Large molecules with many single bonds may have astronomical numbers of local minima, producing the **multiple-minima problem**. For example, consider the  $n$ -residue

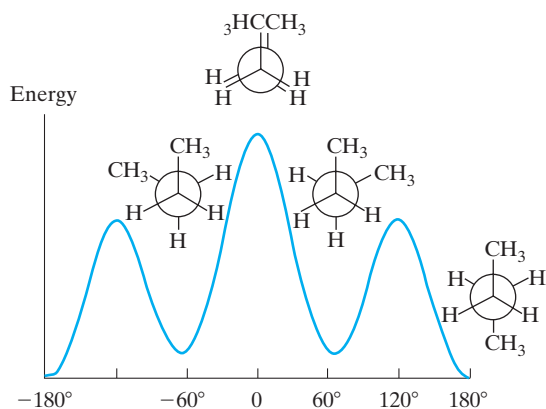
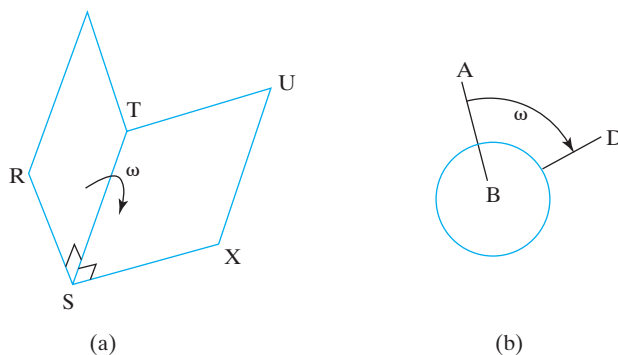


**FIGURE 15.13** Electronic energy (including nuclear repulsion) for  $\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_3$  versus the CCCC dihedral angle.

**FIGURE 15.14** (a) The dihedral angle  $\omega$  between the half planes RST and STU. (b) The dihedral angle  $D(ABCD)$ . Atom C is behind B.



polypeptide  $(-\text{NH}-\text{CHR}-\text{C}(\text{O})-)_n$ . Even if one assumes that each  $\text{OC}-\text{NH}$  dihedral angle is  $180^\circ$  (its most common value) and ignores the conformations of the amino acid side chains R, the polypeptide has  $2n$  adjustable dihedral angles (the torsion angles about the  $\text{NH}-\text{CHR}$  and  $\text{CHR}-\text{C}(\text{O})$  bonds), each of which has three likely minima ( $60^\circ$ ,  $-60^\circ$ , and  $180^\circ$ ). This gives  $3^{2n}$  possible conformations that are local energy minima. For  $n = 40$ , this is  $3^{80} \approx 10^{38}$  conformations to be examined. Methods for searching for low-energy conformations of medium-size and large molecules are discussed in Section 15.11.

The **dihedral angle**  $\omega = D(\text{RSTU})$  in Fig. 15.14a is defined as the angle  $\omega$  between the half plane RST and the half plane STU. More precisely, the dihedral angle is angle RSX, where lines RS and SX are both perpendicular to line ST. In dealing with a molecular dihedral angle  $D(\text{ABCD})$  involving four atoms A, B, C, D, the AB and CD bonds are not likely to be perpendicular to the BC bond, so one must project AB and CD into a plane perpendicular to line BC. This is done by drawing a Newman projection with line BC perpendicular to the plane of the paper (Fig. 15.14b). The range of dihedral angles is chosen to be either  $0^\circ \leq \omega < 360^\circ$  or  $-180^\circ < \omega \leq 180^\circ$ . By definition, a clockwise rotation of the projected front bond AB to bring it to atom D corresponds to a positive dihedral angle  $D(\text{ABCD})$ .

Some references on geometry optimization are H. B. Schlegel in *Yarkony*, Part I, Chapter 8; *Leach*, Chapter 5; H. B. Schlegel, *J. Comput. Chem.*, **24**, 1514 (2003); H. B. Schlegel, *WIREs Comput. Mol. Sci.*, **1**, 790 (2011).

### Geometry Optimization

Many systematic mathematical procedures (algorithms) exist to find a local minimum of a function of several variables. These procedures will find a local minimum in  $U$  in the neighborhood of the initially assumed molecular geometry. The process of finding such a minimum is called **geometry optimization** or **energy minimization**. For a molecule with several conformations, one must repeat the local-minimum search procedure for each possible conformation, so as to locate the global minimum. For large molecules, there may be too many conformations for all of them to be examined. Moreover, the true global-minimum equilibrium geometry might correspond to a highly unconventional structure that the researcher might not think to consider. For example, high-level calculations that use large basis sets and include electron correlation show that, for the vinyl cation, the classical structure (Fig. 15.15a) lies about 4 kcal/mol higher in energy than the true equilibrium structure (Fig. 15.15b), which has a three-center bond [C. Liang et al., *J. Chem. Phys.*, **92**, 3653 (1990); B. T. Psciuk et al., *Theor. Chem. Acc.*, **118**, 75 (2007)]; the infrared spectrum of the vinyl cation also shows the three-center-bond

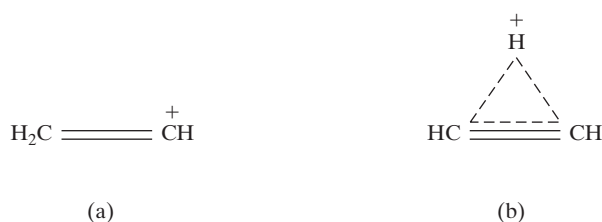


**FIGURE 15.15** (a) Classical structure of the vinyl cation. (b) True equilibrium structure of the vinyl cation.

structure to be more stable [M. W. Crofton et al., *J. Chem. Phys.*, **91**, 5139 (1989)]. The ethyl cation has a similar symmetrical bridged structure [B. Ruscic et al., *J. Chem. Phys.*, **91**, 114 (1989)].

Some procedures to find a local minimum in  $U$  require only repeated calculation of  $U$  at various values of its variables, but these procedures are very inefficient. More efficient procedures require repeated calculation of both  $U$  and its derivatives. The set of  $3N - 6$  first partial derivatives of  $U$  with respect to each of its variables constitutes a vector (in a “space” of  $3N - 6$  dimensions) called the **gradient** of  $U$  (Section 5.2). At a local minimum, the gradient must be zero, meaning that each of the  $3N - 6$  first partial derivatives of  $U$  must be zero. Any point on the PES where the gradient is zero is called a **stationary** (or *critical*) **point**. A stationary point on a PES may be a minimum, a maximum, or a saddle point.

Analytical calculation of the gradient is the key to efficient geometry optimization. The SCF MO energy expression is Eq. (14.44), and its derivatives with respect to the nuclear coordinates would seem to involve the derivatives of the  $H_{rs}^{\text{core}}$  and  $(rs|tu)$  integrals (which occur in  $\epsilon_i$ ), the derivatives of  $V_{NN}$ , and the derivatives of the SCF coefficients  $c_{si}$  (which occur in  $P_{rs}$ ). However, the terms involving the derivatives of the  $c_{si}$ ’s turn out to add up to zero, leaving only the derivatives of the integrals and of  $V_{NN}$ . The derivatives of the integrals are readily calculated, since the derivative of a Gaussian-type function with respect to a nuclear coordinate is another GTF. The derivatives of  $V_{NN}$  are trivial. Thus an analytic formula for the gradient of the ab initio SCF energy is known [see *Hehre* et al., Section 3.3.3; P. Pulay, *Adv. Chem. Phys.*, **69**, 241 (1987); P. Pulay in *Yarkony*, Part II, Chapter 9]. Once the SCF energy  $U$  and wave function have been found for some chosen geometry, the time needed to analytically calculate the energy gradient is roughly equal to the time needed to do the SCF wave function and energy calculation. (To calculate the gradient numerically requires varying the  $3N - 6$  nuclear coordinates one at a time by a small amount, repeating the SCF calculation for each new geometry to get  $U$  for that geometry, and estimating each derivative as the ratio of the change in  $U$  to the change in the coordinate. Numerical evaluation of the gradient thus takes about  $3N - 6$  times as long as analytic evaluation of the gradient.)

As well as using the energy gradient (the  $3N - 6$  first partial derivatives  $\partial U/\partial q_1, \partial U/\partial q_2, \dots$ ), some energy minimization methods also use the second derivatives of  $U$ . The set of second derivatives  $\partial^2 U/\partial q_1^2, \partial^2 U/\partial q_1 \partial q_2, \partial^2 U/\partial q_2 \partial q_1, \partial^2 U/\partial q_2^2, \dots$ , when arranged in a square array form a matrix called the **Hessian** or the **force-constant matrix** [since the second derivatives of  $U$  are force constants; see the equation after (13.21)]. An efficient way to find a local minimum of a function of several variables is the **Newton** (or **Newton–Raphson**) method, which approximates the function by a Taylor-series expansion that is terminated after the quadratic terms, and uses accurately evaluated first and second partial derivatives of the function (which occur in the linear and quadratic terms of the Taylor series). Because analytic calculation of the second derivatives in an ab initio SCF calculation is very costly in computer time, ab initio SCF geometry optimizations commonly use a modification of the Newton procedure,

called the **quasi-Newton** (or *quasi-Newton–Raphson* or *variable metric*) method. In the quasi-Newton method, one does not calculate the Hessian directly, but instead starts with an estimate (or guess) for the Hessian and gradually improves (updates) this estimate using gradient information calculated at each step in the optimization cycle.

To optimize the geometry, one starts with a guess for the equilibrium structure. The guessed structure is based on typical values for bond lengths, bond angles estimated from a method such as the VSEPR method, and dihedral angles guessed based on experience with similar compounds. [One can use a model-builder computer program (Section 15.15) to get an initial structure guess.]

Some typical bond lengths are listed in **Table 15.5**, where X<sub>n</sub> denotes an atom X bonded to n neighbors and *Car* is an aromatic carbon.

The VSEPR (valence-shell electron-pair repulsion) method predicts the bond angles at an atom based on the number of valence electron pairs around that atom in the Lewis electron-dot formula of the molecule. The VSEPR method predicts 180° angles for two pairs, 120° angles for three pairs, 109½° (tetrahedral) angles for four pairs, and 90° angles (octahedral arrangement of pairs) for six pairs. For example, the Lewis structure of H<sub>2</sub>O shows four pairs around the O atom, which indicates a tetrahedral 109½° bond angle. Because lone pairs occupy a larger volume of space than bonded pairs, we expect the angle to be a little less than 109½°. The observed angle is 104½°. The VSEPR method counts a double or triple bond as only one pair. Thus, 120° angles are predicted at each carbon in H<sub>2</sub>C=CH<sub>2</sub>, but because the double bond contains two pairs, we expect the HCH angle to be somewhat less than 120°. For further details on the VSEPR method, see any general chemistry textbook or R. J. Gillespie and I. Hargittai, *The VSEPR Model of Molecular Geometry*, Prentice Hall, 1991.

Some rules for predicting dihedral angles in acyclic organic compounds follow [adapted from J. A. Pople and M. Gordon, *J. Am. Chem. Soc.*, **89**, 4253 (1967)].

1. The conformation about a bond connecting two atoms each with tetrahedral bond angles is usually staggered.
2. For an atom A with tetrahedral angles bonded to an atom B with trigonal (120°) bond angles, (a) one of the non-B atoms bonded to A lies in the same plane as B and the atoms bonded to B; (b) the lowest energy conformer usually has the double bond on B eclipsing a single bond to A.
3. When two bonded atoms A and B both have trigonal bond angles, all atoms bonded to A and B lie in the same plane.

For example, rule 2 tells us that one of the methyl hydrogens in CH<sub>3</sub>—CH=CH<sub>2</sub> is in the same plane as the CH=CH<sub>2</sub> part of the molecule and eclipses the double bond, and further tells us that formic acid, HC(O)OH, is planar with the lowest-energy conformer having the OH hydrogen eclipsing the double bond.

**TABLE 15.5** Some Typical Bond Lengths in Angstroms<sup>a</sup>

|       |       |       |       |        |       |       |       |       |
|-------|-------|-------|-------|--------|-------|-------|-------|-------|
| C4—H  | C3—H  | C2—H  | C4—C4 | C4—C3  | C4—C2 | C3—C3 | C3—C2 | C2—C2 |
| 1.09  | 1.08  | 1.06  | 1.54  | 1.52   | 1.46  | 1.46  | 1.45  | 1.38  |
| C3=C3 | C3=C2 | C2=C2 | C2≡C2 | CarCar | O2—O2 | O2—H  | O2—C4 | O2—C3 |
| 1.34  | 1.31  | 1.28  | 1.20  | 1.40   | 1.48  | 0.96  | 1.43  | 1.36  |
| O2—C2 | O1=C3 | O1=C2 | N3—H  | N2—H   | N3—C4 | N2—C4 | N3—N3 |       |
| 1.36  | 1.22  | 1.16  | 1.01  | 0.99   | 1.47  | 1.47  | 1.45  |       |

<sup>a</sup>Excerpted from J. A. Pople and M. Gordon, *J. Am. Chem. Soc.*, **89**, 4253 (1967).

After guessing the geometry, one searches for the minimum nearest the initially assumed geometry. One chooses a basis set and uses the SCF MO (or some other) method to approximately solve the electronic Schrödinger equation to find  $U$  and its gradient at the guessed initial geometry. Using the calculated values of  $U$  and  $\nabla U$  (and perhaps information about the Hessian), the geometry optimization program changes the  $3N - 6$  nuclear coordinates to a new set that is likely to be closer to a minimum than the initial set, and the SCF  $U$  and  $\nabla U$  are calculated at this new structure. Using the results of the new calculation, a further improved set of nuclear coordinates is calculated, and the SCF calculation is repeated at the new geometry. The process is repeated until  $\nabla U$  differs negligibly from zero, indicating that a minimum (at which  $\nabla U$  is zero) may have been found. Typically, about  $3N - 6$  to  $2(3N - 6)$  repetitions of the SCF and gradient calculations are needed to find a minimum. The availability of analytical gradients makes possible the efficient determination of the ab initio equilibrium geometry of small- and medium-size molecules, and the introduction of analytical gradients into ab initio calculations (by Pulay in 1969) has been called a “revolution” [L. Schäfer, *J. Mol. Struct.*, **100**, 51 (1983)].

Common methods of geometry optimization can converge to a stationary point that is not a minimum, but is a saddle point. For example, if one does a HF/6-31G\* calculation on  $\text{NH}_3$  starting with a planar geometry, the geometry optimizer in the program *Gaussian 09* converges to a planar geometry with 0.99 Å bond lengths and  $120^\circ$  angles. This geometry is a maximum with respect to motion of the nitrogen atom in the direction perpendicular to the molecular plane but is a minimum with respect to the remaining coordinates. Hence this stationary point is a first-order saddle point. *To be sure that one has found a minimum and not a saddle point, it is essential to test the nature of the stationary point found by the geometry optimization.* This is done by doing a vibrational-frequency calculation (Section 15.12) at the geometry found. For a true minimum, all the calculated frequencies will be real. For a first-order saddle point, one calculated frequency will be imaginary. Sad to say, there are optimized geometries reported in the literature that are saddle points rather than minima.

As an example of geometry optimization, a HF/3-21G geometry optimization of the formic acid molecule  $\text{HC(O)OH}$  using the *Gaussian 09* program converged in three steps as shown in Table 15.6. The initial geometry uses bond distances from Table 15.5 and bond angles given by VSEPR and assumes the molecule is planar, as predicted by the preceding dihedral-angle rules. A conformation with the OH hydrogen eclipsing the carbonyl oxygen was assumed. (For the other likely conformation, see Prob. 15.57.) The final optimized geometry is that at step 3, since the predicted coordinate changes from step 3 to step 4 are all less than the program’s cutoff for convergence and the magnitudes of the gradient components at the step 3 geometry are all less than the cutoff value. The maximum bond-length and bond-angle changes from step 3 to step 4 would have been 0.0003 Å and  $0.06^\circ$ . A frequency calculation at the step 3 geometry gives all real

**TABLE 15.6 HF/3-21G Geometry Optimization of a Formic Acid Conformer<sup>a</sup>**

|          | R(CH) | R(C=O) | R(CO) | R(OH)  | $\angle(\text{HC=O})$ | $\angle(\text{OCO})$ | $\angle(\text{COH})$ |
|----------|-------|--------|-------|--------|-----------------------|----------------------|----------------------|
| initial  | 1.080 | 1.220  | 1.360 | 0.960  | $120.0^\circ$         | $120.0^\circ$        | $109.5^\circ$        |
| step 1   | 1.070 | 1.205  | 1.347 | 0.970  | 124.8                 | 125.1                | 110.5                |
| step 2   | 1.074 | 1.197  | 1.349 | 0.972  | 126.2                 | 124.4                | 113.5                |
| step 3   | 1.074 | 1.198  | 1.350 | 0.9696 | 125.78                | 124.69               | 112.76               |
| (step 4) | 1.074 | 1.198  | 1.350 | 0.9695 | 125.73                | 124.64               | 112.69               |

<sup>a</sup>In this conformer, the OH hydrogen eclipses the carbonyl oxygen. Bond lengths are in angstroms.

vibrational frequencies, confirming that this geometry is a minimum and not a saddle point. The energy  $U$  and the magnitude  $|\nabla U|$  of the gradient at the steps of this optimization are as follows:

|                                            | initial     | step 1      | step 2      | step 3      |
|--------------------------------------------|-------------|-------------|-------------|-------------|
| $U/\text{hartrees}$                        | -187.694797 | -187.699879 | -187.700158 | -187.700199 |
| $ \nabla U /(\text{hartrees}/\text{bohr})$ | 0.0723      | 0.0217      | 0.0061      | 0.0007      |

Note the decreasing energy and the decreasing magnitude of the gradient as the minimum is approached.

The choice of coordinates used in the search affects the convergence rate of the optimization. One choice is the **Cartesian coordinates** of the nuclei. Another choice is to use bond distances, bond angles, and dihedral angles, which constitute the **primitive internal coordinates**. (Internal coordinates describe molecular vibrations but not translations or rotations.) For nonrigid molecules containing more than 30 atoms, certain linear combinations of the primitive internal coordinates, called *delocalized internal coordinates*, perform much better than Cartesian coordinates. For example, a semiempirical SCF optimization of the antibiotic jawsamycin ( $\text{C}_{32}\text{H}_{13}\text{N}_3\text{O}_6$ ,  $3N - 6 = 246$ ) required 409 cycles of coordinate changes using Cartesian coordinates but only 71 cycles using delocalized internal coordinates [J. Baker et al., *J. Chem. Phys.*, **105**, 192 (1996); **110**, 4986 (1999)]. Several other choices of internal coordinates are in use [see V. Bakken and T. Helgaker, *J. Chem. Phys.*, **117**, 9160 (2002); K. Németh and M. Challacombe, *ibid.*, **121**, 2877 (2004)].

Ab initio SCF MO single-point calculations on molecules with several hundred atoms are now feasible with small basis sets. However, ab initio geometry optimization of such large molecules is not yet routinely possible, unless the molecule is highly symmetrical.

### Notation

We shall use the ideas of an  $n$ -dimensional vector space—Section 5.2. Let  $\mathbf{q}$  be the  $(3N - 6)$ -dimensional vector whose components are the nuclear coordinates that define the molecular geometry. [If Cartesian coordinates are used in the optimization, we deal with a  $3N$ -dimensional vector; if internal coordinates are used, we have a  $(3N - 6)$ -dimensional vector.] Thus the molecular geometry corresponds to a point in  $(3N - 6)$ -dimensional space. Let  $\mathbf{q}_1$  be the initially assumed geometry, and let  $\mathbf{q}_2, \mathbf{q}_3, \dots, \mathbf{q}_k, \mathbf{q}_{k+1}, \dots$  be the geometries generated by the optimization procedure. The geometry-optimization procedure produces a series of steps. The  $k$ th step is defined by the vector  $\Delta\mathbf{q}_k$  that goes from  $\mathbf{q}_k$  to  $\mathbf{q}_{k+1}$ ; thus  $\Delta\mathbf{q}_k = \mathbf{q}_{k+1} - \mathbf{q}_k$ . The components of  $\Delta\mathbf{q}_k$  are the changes in each nuclear coordinate for step  $k$ . The *length* and *direction* of the  $k$ th step are the length and direction of the vector  $\Delta\mathbf{q}_k$ .

### The quasi-Newton Method

To illustrate the quasi-Newton method, we shall pretend that  $U$  is a function of only two variables  $X$  and  $Y$ . Let  $X_1$  and  $Y_1$  denote the initially assumed molecular geometry. If we neglect terms higher than quadratic, the Taylor series for a function of two variables is (Sokolnikoff and Redheffer, p. 336)

$$\begin{aligned}
 U(X, Y) \approx U(X_1, Y_1) &+ \left. \frac{\partial U}{\partial X} \right|_{X_1, Y_1} (X - X_1) + \left. \frac{\partial U}{\partial Y} \right|_{X_1, Y_1} (Y - Y_1) + \frac{1}{2} \left. \frac{\partial^2 U}{\partial X^2} \right|_{X_1, Y_1} (X - X_1)^2 \\
 &+ \left. \frac{\partial^2 U}{\partial X \partial Y} \right|_{X_1, Y_1} (X - X_1)(Y - Y_1) + \frac{1}{2} \left. \frac{\partial^2 U}{\partial Y^2} \right|_{X_1, Y_1} (Y - Y_1)^2 \quad (15.67)
 \end{aligned}$$

Let  $U_X \equiv \partial U/\partial X$ ,  $U_Y \equiv \partial U/\partial Y$ ,  $U_{XX} \equiv \partial^2 U/\partial X^2$ ,  $U_{XY} \equiv \partial^2 U/\partial X \partial Y$ , and  $U_{YY} \equiv \partial^2 U/\partial Y^2$ ; let the subscript 1 denote evaluation at the point  $(X_1, Y_1)$ . Then

$$U(X, Y) \approx U_1 + U_{X,1}(X - X_1) + U_{Y,1}(Y - Y_1) + \frac{1}{2}U_{XX,1}(X - X_1)^2 + U_{XY,1}(X - X_1)(Y - Y_1) + \frac{1}{2}U_{YY,1}(Y - Y_1)^2 \quad (15.68)$$

Just as the harmonic-oscillator approximation to a diatomic molecule's  $U(R)$  function works well in the region near  $R_e$  (Fig 4.6), the quadratic approximation (15.68) works well in the region near a minimum.

If  $U$  were accurately a quadratic function of the coordinates in the region near  $(X_1, Y_1)$ , then the second partial derivatives (the elements of the Hessian matrix) would be constants in this region, and the subscript 1 on the second partials would be unnecessary. Accurate ab initio SCF calculation of the second derivatives is very time-consuming, so one usually uses a quasi-Newton method, meaning that one starts with an approximation for the Hessian and improves this approximation as the geometry optimization proceeds. We therefore write

$$U(X, Y) \approx U_1 + U_{X,1}(X - X_1) + U_{Y,1}(Y - Y_1) + \frac{1}{2}U_{XX}^{(1)}(X - X_1)^2 + U_{XY}^{(1)}(X - X_1)(Y - Y_1) + \frac{1}{2}U_{YY}^{(1)}(Y - Y_1)^2 \quad (15.69)$$

where the superscript<sup>(1)</sup> denotes our first approximation to the Hessian matrix elements at (or near) the equilibrium geometry.

How do we get the initial guesses for the second derivatives of  $U$ ? These derivatives are force constants, and force constants for stretching or bending a particular kind of bond length or angle are roughly constant from molecule to molecule. Thus if we are dealing with a compound containing an  $\text{H}-\text{C}=\text{O}$  group, we use known typical force constants for stretching the  $\text{H}-\text{C}$  and  $\text{C}=\text{O}$  bonds and for bending the  $\text{HCO}$  angle to help construct an initial estimate of the Hessian. [This approach is closely related to the molecular-mechanics method (Section 17.5).] A procedure for estimating the initial Hessian based on typical force constants is built into most programs that do geometry optimization. An alternative would be to get the initial Hessian from a semiempirical calculation (Chapter 17), which is much faster than an ab initio calculation.

Partial differentiation of (15.69) with respect to  $X$  and with respect to  $Y$  gives

$$\begin{aligned} U_X(X, Y) &\approx U_{X,1} + U_{XX}^{(1)}(X - X_1) + U_{XY}^{(1)}(Y - Y_1) \\ U_Y(X, Y) &\approx U_{Y,1} + U_{XY}^{(1)}(X - X_1) + U_{YY}^{(1)}(Y - Y_1) \end{aligned} \quad (15.70)$$

At a minimum,  $U_X(X, Y)$  and  $U_Y(X, Y)$  are zero. Let  $(X_2', Y_2')$  denote the point where the estimated first derivatives  $U_X$  and  $U_Y$  (the gradient components) on the left sides of (15.70) vanish. At  $(X, Y) = (X_2', Y_2')$ , Eq. (15.70) becomes

$$\begin{aligned} 0 &= U_{X,1} + U_{XX}^{(1)}(X_2' - X_1) + U_{XY}^{(1)}(Y_2' - Y_1) \\ 0 &= U_{Y,1} + U_{XY}^{(1)}(X_2' - X_1) + U_{YY}^{(1)}(Y_2' - Y_1) \end{aligned} \quad (15.71)$$

Solving for  $X_2'$  and  $Y_2'$ , we get

$$X_2' = X_1 + \frac{U_{XY}^{(1)}U_{Y,1} - U_{YY}^{(1)}U_{X,1}}{U_{XX}^{(1)}U_{YY}^{(1)} - (U_{XY}^{(1)})^2}, \quad Y_2' = Y_1 + \frac{U_{XY}^{(1)}U_{X,1} - U_{XX}^{(1)}U_{Y,1}}{U_{XX}^{(1)}U_{YY}^{(1)} - (U_{XY}^{(1)})^2} \quad (15.72)$$

Starting at the initially guessed geometry  $(X_1, Y_1)$ , we have used the calculated gradient at point 1 and the initial guess for the Hessian to find point 2'. The step from point 1 to point 2' calculated from (15.72) is called a *Newton (-Raphson) step*. If  $U$  were truly a quadratic function in the region we are working in and if we had accurate values for the Hessian matrix elements, the formula (15.72) would give us the minimum

in  $U$  in a single step. In actuality,  $(X_2', Y_2')$  is only an approximation to the point that minimizes  $U$ .

We now use the ab initio SCF MO method (or some other method) to calculate  $U$  and its gradient at  $(X_2', Y_2')$ .

We could now use  $(X_2', Y_2')$  as the new starting geometry for the next cycle of the geometry optimization, but faster convergence is obtained if instead we take the new starting point as  $X_2 = X_1 + \alpha(X_2' - X_1)$ ,  $Y_2 = Y_1 + \alpha(Y_2' - Y_1)$ , where the value of  $\alpha$  is found as follows. One expresses  $U$  as a polynomial (typically a cubic or quartic) whose coefficients are determined so that the  $U$  polynomial will have the values that were calculated for  $U$  at  $(X_1, Y_1)$  and at  $(X_2', Y_2')$ , and the gradient of  $U$  will have the calculated gradient values at these two geometries. One then varies  $\alpha$  to minimize the  $U$  polynomial, thereby giving the new predicted geometry  $(X_2, Y_2)$ . This is an example of a **line search**. By varying  $\alpha$ , we are searching along a line in  $(3N - 6)$ -dimensional space that joins  $(X_1, Y_1)$  and  $(X_2', Y_2')$ .

Having obtained the new geometry  $(X_2, Y_2)$ , we could now do an SCF calculation of  $U$  and its gradient at  $(X_2, Y_2)$ , but it is accurate enough to just use the interpolated values of these quantities found from the  $U$  polynomial that was fitted to the data at points 1 and 2'.

One now uses the values of the gradient of  $U$  at points 1 and 2 to improve (update) the estimate of the Hessian by requiring that the improved Hessian satisfy Eq. (15.70) for points 1 and 2. Using a superscript <sup>(2)</sup> to denote the improved Hessian matrix elements, we require that

$$\begin{aligned} U_{X,2} &= U_{X,1} + U_{XX}^{(2)}(X_2 - X_1) + U_{XY}^{(2)}(Y_2 - Y_1) \\ U_{Y,2} &= U_{Y,1} + U_{XY}^{(2)}(X_2 - X_1) + U_{YY}^{(2)}(Y_2 - Y_1) \end{aligned} \quad (15.73)$$

There are three Hessian matrix elements to be solved for, but only two equations to be satisfied, so there is not a unique solution for the  $U^{(2)}$ 's, and several recipes have been proposed to find improved  $U^{(2)}$ 's that satisfy (15.73). One widely used recipe is the Broyden, Fletcher, Goldfarb, Shanno (BFGS) procedure (*Leach*, Section 5.6).

Having improved the Hessian, we now use (15.72) with the  $U^{(1)}$ 's replaced by the  $U^{(2)}$ 's and with point 1 replaced by point 2 to calculate the new coordinates  $(X_3', Y_3')$ . We then check for convergence by seeing if the absolute values of the predicted coordinate changes  $X_3' - X_2$  and  $Y_3' - Y_2$  are both less than some tiny fixed amount and if the gradient components  $|U_{X,2}|$  and  $|U_{Y,2}|$  are both less than some tiny amount. (The gradient must vanish at a minimum.) If all these conditions are met, the optimization is finished and the predicted geometry is point 2. If convergence has not been achieved, we calculate  $U$  and  $\nabla U$  at point 3', do a line search between points 2 and 3' to locate point 3; and so on.

For a function of many variables, the best way to write the Taylor series and the geometry-search equations is using matrices.

Early in the quasi-Newton procedure when one is not very close to a minimum, the procedure may well predict large coordinate changes for which the quadratic approximation to the PES may be quite inaccurate and the predicted quasi-Newton step might make things worse rather than better. To avoid this problem, one imposes a *trust radius*. When the length of a predicted step exceeds the trust radius, the coordinate changes are reduced by a scale factor; also, the direction of the step may be varied from the quasi-Newton prediction using some other search procedure.

In the quasi-Newton method, the next geometry is obtained from the Newton formula (15.72) plus a line search. A commonly used alternative to the quasi-Newton method is to calculate the next set of nuclear coordinates by a modified form of (15.72) in which the current coordinates  $X_1, Y_1$  are replaced by linear combinations of the current coordinates and the coordinates in all the previous search steps, and the current gradient components

are replaced by similar linear combinations of current and previous gradient components. The coefficients in the linear combinations are chosen so as to minimize the distance in the  $3N - 6$  dimensional space from the point that is the linear combination of coordinates to the next predicted geometry point. No line search is needed. The Hessian may or may not be updated. This procedure is the **GDIIS** (geometry by direct inversion in the iterative subspace) **method** [P. Császár and P. Pulay, *J. Mol. Struct.*, **114**, 31 (1984); O. Farkas and H. B. Schlegel, *Phys. Chem. Chem. Phys.*, **4**, 11 (2002)].

In the molecular-mechanics method (Section 17.5), analytic evaluation of the second derivatives of  $U$  is rapid, so (provided the molecule is not very large) instead of the quasi-Newton method, one can use the Newton (–Raphson) method, in which the Hessian is accurately calculated instead of being estimated. The molecular-mechanics method allows geometry optimization for molecules containing thousands of atoms. For such large molecules, the Newton method is too computationally demanding, since one must deal with a large Hessian matrix. For very large molecules, molecular-mechanics geometry optimizations often use a modification of the Newton–Raphson method called the **block-diagonal Newton–Raphson method**. Here, one makes the approximation that  $\partial^2 U / \partial q_i \partial q_j = 0$  whenever  $q_i$  and  $q_j$  are Cartesian coordinates of different atoms. This approximation puts the Hessian matrix in block-diagonal form, where each block is  $3 \times 3$  and contains 9 second partial derivatives that involve only the coordinates of a particular atom. This allows us to deal with the atoms one at a time. For a 1000-atom molecule, instead of having to deal with a Hessian containing  $3000^2 = 9 \times 10^6$  elements (or  $2994^2$  elements after vibrations and rotations are removed), we deal with 1000 matrices, each containing only  $3^2 = 9$  elements.

An alternative when the size of the molecule prevents use of the quasi-Newton or Newton–Raphson methods is to choose an optimization method that uses only the gradient and not the Hessian. Two such methods are the steepest-descent method and the conjugate-gradient method.

### The Steepest-Descent and Conjugate-Gradient Methods

In the **steepest-descent method**, one begins by calculating  $U$  and  $\nabla U$  at the initially assumed geometry. Let these quantities be  $U_1$  and  $\nabla U_1$ . Recall (Section 5.2) that the vector  $\nabla U$  points in the direction of greatest rate of increase in  $U$ . In the steepest-descent method, each search step is taken in the direction for which  $U$  decreases the fastest, which means that the first step is in the direction of  $-\nabla U_1$ . (This direction is perpendicular to the contour surface of constant  $U$  that goes through point 1.) The size of the step is determined by a line search, as follows. One calculates  $U$  at several points along the  $-\nabla U_1$  direction, fits a polynomial to the calculated  $U$  values on the line, and finds the minimum of the fitted polynomial, thereby giving point 2. One then calculates  $U_2$  and  $\nabla U_2$  and does a line search in the direction of  $-\nabla U_2$ . One continues until the gradient and predicted step size have become negligibly small. The steepest-descent method can be very inefficient near the end of the search where  $\nabla U$  is small, and so is used only at the beginning of the search when one is not close to a minimum point and  $\nabla U$  is large. One then switches to another search method when  $\nabla U$  becomes small.

An improvement on the steepest-descent method is the **conjugate-gradient method**. Here, the first step is the same as in the steepest-descent method, so  $\mathbf{q}_2 = \mathbf{q}_1 - \lambda_1 \nabla U_1$ , where  $\lambda_1$  is found from a line search. The direction of each subsequent step  $k$  is defined by a vector  $\mathbf{d}_k$  (where  $k = 2, 3, \dots$ ) that is a linear combination of the negative gradient  $-\nabla U_k$  and the preceding search direction. The explicit formulas for the conjugate-gradient method are

$$\mathbf{q}_{k+1} = \mathbf{q}_k + \lambda_k \mathbf{d}_k$$

$$\mathbf{d}_1 \equiv -\nabla U_1 \quad \text{and} \quad \mathbf{d}_k \equiv -\nabla U_k + \beta_k \mathbf{d}_{k-1} \quad \text{for } k > 1$$

The constant  $\lambda_k$  is found by a line search that minimizes  $U$  in the direction of  $\mathbf{d}_k$ . In the Fletcher–Reeves version of the conjugate-gradient method,  $\beta_k$  is calculated from the formula

$$\beta_k \equiv (\nabla U_k \cdot \nabla U_k) / (\nabla U_{k-1} \cdot \nabla U_{k-1})$$

(An alternative formula for  $\beta_k$  is the Polak–Ribiere formula; see *Leach*, Section 5.4.4.) The idea of the conjugate-gradient method (which really should be called the conjugate-direction method) is to choose each new step in a direction that is conjugate to the directions used in the previous steps (where the word “conjugate” has a certain technical meaning that will not be defined here), so as to avoid undoing the minimization work done in previous steps.

### The Truncated Newton Method

In the Newton and quasi-Newton methods, one solves a set of linear equations like (15.71) to find each Newton–Raphson step. For large molecules, repeated solution of these linear equations is time-consuming. Early in the search, when one is not very close to the minimum, the Newton step direction is not expected to be that accurate, and it is a waste of time to solve for this direction accurately. The **truncated Newton (TN) method** therefore solves these linear equations only approximately. The method is programmed so that the accuracy with which the linear equations are solved increases as the gradient decreases and one comes closer to the minimum. In the TN method the linear equations are often solved with a conjugate-gradient procedure, giving a procedure labeled TNCG. The TNCG method is often used in molecular-mechanics geometry optimizations.

## 15.11 Conformational Searching

Large molecules may have huge numbers of conformations that are energy minima. One is usually interested in finding not only the lowest-energy conformation (the global minimum on the PES), but also all minima whose energies are low enough so that these conformations have significant populations at room temperature. The biologically active conformer of a biomolecule might not be the global-minimum conformation. Because of entropy effects, the global minimum might not be the most populated conformer at room temperature. Currently, no method exists that is guaranteed to find the global minimum and all low-lying minima of a large flexible molecule. Many methods of conformational searching exist. Because of the many possible conformers, the geometry optimization part of a conformational search for a large molecule is usually done using the molecular-mechanics method (Sections 15.1 and 17.5), rather than a quantum-mechanical calculation. (For a fuller discussion, see *Leach*, Chapter 9.)

In the **systematic (or grid) search method**, one uses a computer to systematically increment each dihedral angle involving rotation about a single bond by a fixed amount  $\Delta\theta$  until all possible combinations of dihedral angles for the chosen  $\Delta\theta$  have been generated. Typical values for  $\Delta\theta$  are  $30^\circ$ ,  $60^\circ$ , or  $120^\circ$ . The larger the value of  $\Delta\theta$ , the fewer the number of possibilities that have to be examined and the more likely it is that minima might be missed. When each new set of dihedral angles is generated, one first checks that the configuration produced does not have any nonbonded atoms too close to each other (as determined by the van der Waals radii of the atoms); this is called a **bump check**. If the conformation generated passes the bump check, one uses a geometry-optimization procedure to find the nearest energy minimum on the PES. When the nearest minimum has been found, one checks that it differs from any previously found minima. Special procedures are needed to apply a systematic search to ring compounds. Systematic searches are limited to molecules with no more than about 15 or 20 dihedral angles.

In the **random** (or **stochastic** or **Monte Carlo**) **search method**, one starts with a stable conformer and generates new initial configurations either by randomly changing the values of randomly chosen dihedral angles or by adding small random amounts to the Cartesian coordinates of each atom. This is followed by a bump check and energy minimization. (The term **Monte Carlo** denotes random sampling of points, rather than taking a regular grid of points. Monte Carlo is a European gambling resort where the laws of probability are continually tested.)

The **distance-geometry** search method describes the molecule by a **distance matrix** whose elements  $d_{ij}$  are the distances between all possible pairs of atoms  $i$  and  $j$ . One begins by assigning a minimum and maximum permitted value to each internuclear distance. For two atoms bonded to each other (1,2 atoms), the minimum and maximum permitted values are usually set equal to the typical bond length for the two atoms and the kind of bond (Table 15.5). The distance between two atoms A and C each bonded to the same atom B (1,3 atoms) is usually set equal to the value determined from the assigned A—B and B—C bond distances and the ABC bond angle, whose value can be taken as a standard value ( $109\frac{1}{2}^\circ$  for a tetrahedral angle, and so on). (Instead of using standard bond distances and bond angles to fix the 1,2 and 1,3 distances, one can use the bond distances and bond angles in a single conformer that has been energy minimized.) For two atoms A and D separated by three bonds (1,4 atoms, A—B—C—D) the minimum allowed A to D distance is the distance when the ABCD dihedral angle is 0 and the maximum allowed distance is the value corresponding to a  $180^\circ$  dihedral angle.

For a 1, $n$  pair of atoms with  $n > 4$ , the minimum allowed distance is initially set equal to the sum of the van der Waals radii of the atoms, and the maximum allowed distance is temporarily set equal to some large number (say, 100 Å). These maximum allowed distances are then reduced using the **triangle inequality**  $d_{AE} \leq d_{AG} + d_{GE}$ , where A, E, and G are any three nuclei. (To see the validity of this relation, just join A, E, and G to form a triangle.) One repeatedly examines all possible sets of three nuclei and lowers each AE maximum-allowed distance that does not satisfy  $u_{AE} \leq u_{AG} + u_{GE}$ , where  $u$  denotes the upper bound (maximum allowed value). The preceding triangle inequality can be written as  $d_{AG} \geq d_{AE} - d_{GE}$ . One therefore repeatedly examines all trios of nuclei and raises any lower-bound distances  $l_{AG}$  that do not satisfy  $l_{AG} \geq l_{AE} - u_{GE}$ .

Once the final values of the minimum and maximum allowed distances have been arrived at, the distance-geometry method assigns each internuclear distance a random value that lies in its permitted range. These assigned internuclear distances may well not correspond to a possible arrangement of the atoms in three-dimensional space, so the distance-geometry method then uses mathematical procedures that produce a set of nuclear Cartesian coordinates for which the internuclear distances lie as close as possible to the randomly chosen values and that lie (as far as is possible) within the permitted ranges. Starting from these nuclear coordinates, one then does an energy minimization to find a conformer. (The equations of distance geometry, which find nuclear Cartesian coordinates from internuclear distances, aid in converting internuclear distances found by two-dimensional NMR spectroscopy of proteins into protein structures.)

The **genetic algorithm** (GA) method uses procedures analogous to mating, mutation, and survival of the fittest in living organisms. (The GA method is an example of an *evolutionary algorithm* method.) In a GA conformational search, each molecular dihedral angle is expressed as a string of  $n$  zeros and ones (*bits*—binary digits), where  $n$  is typically in the range 6 to 12. For example, with  $n = 9$ , the string 011010001 is the binary number  $0(2^8) + 1(2^7) + 1(2^6) + 0(2^5) + 1(2^4) + 0(2^3) + 0(2^2) + 0(2^1) + 1(2^0) = 209$ , and the dihedral angle is defined as  $(209/2^9)360^\circ = 147^\circ$ . (For technical reasons, the GA method often uses a different procedure, called Gray coding, to represent each angle as a binary string. In Gray coding, the binary representations of the successive decimal

integers  $m$  and  $m + 1$  differ by a single bit change. See R. Judson in K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry*, Vol. 10, Wiley, 1997, Chapter 1, which is a good review article on the GA method.) The set of single-bond dihedral angles in a molecule is encoded by stringing together the strings that encode each dihedral angle, thereby forming a “chromosome” containing  $nd$  bits, where  $d$  is the number of single-bond dihedral angles. A chromosome encodes the conformation of a molecule.

The GA method typically begins by generating a set of about 100 first-generation chromosomes, where each such chromosome is formed by setting each of its  $nd$  bits randomly equal to 0 or 1. For each chromosome, one does an energy calculation using the molecular-mechanics method to find the chromosome’s molecular electronic energy  $U$  (really its steric energy; see Section 17.5). The lowest-energy chromosome is ranked as the fittest, the second lowest is the second fittest, and so on.

To form the next generation, one first takes the fittest 10% of chromosomes and moves them into the next generation; the fittest chromosome is moved unchanged, and the rest of the fittest 10% are each subjected to a small probability of a mutation (a random change in a bit). The remaining 90% of the second generation is formed by mating. Mating is done by first picking two chromosomes at random from the breeding pool, which typically consists of the fittest 40% of the first generation. One then forms two “children” of the chosen pair by replacing the first  $m$  bits of each chosen chromosome with the corresponding first  $m$  bits of the other chromosome, where  $m$  is a randomly chosen number in the range 1 to  $nd - 1$ . In addition to interchanging the first part of the two parent chromosomes, a small probability for mutation is included. Enough matings are done to make the size of the second generation equal that of the first. Energy evaluations and rankings are then done for the second-generation chromosomes, and the third generation is formed from the second the same way the second was formed from the first.

One continues for typically 100 generations. Then the members of the last generation are sorted into groups, where each group has a similar set of dihedral angles, and the lowest-energy member of each group is energy minimized (Section 15.10). The process can be repeated by choosing a new set of random first-generation chromosomes.

In the **molecular dynamics** search method, one begins with a conformation that is a minimum and then assigns to each atom a set of velocity components  $v_x, v_y, v_z$  that are randomly chosen from a Maxwell distribution at an elevated temperature (typically 500 K or 1000 K). The initial position and velocity of each atom are therefore known. One then applies Newton’s second law of motion to each atom, where each component of the force on atom  $i$  is calculated as  $F_{x,i} = -\partial U / \partial x_i$ , where  $U$  is obtained from a molecular-mechanics force field (Section 17.5). One numerically integrates Newton’s second law to get the position of each atom at the times  $\Delta t, 2\Delta t, 3\Delta t, \dots$ , where the time interval  $\Delta t$  is typically  $10^{-15}$  s and one typically follows the atomic motions for  $10^{-9}$  s. At equal time intervals, one samples configurations generated by the atomic motions (typically  $10^3$  to  $10^4$  configurations are taken) and each sampled configuration is subjected to energy minimization, so as to find a conformation. The elevated temperature used allows the molecule to climb over potential energy barriers to reach new regions of the PES that may contain lower-energy minima than the current region.

In the **Metropolis Monte Carlo** search method, one assigns the molecule an elevated temperature  $T$  (typically 1000 K). Starting with an initial conformation of the molecule, one randomly changes one (or a few) randomly chosen dihedral angles to give a new conformation whose energy is evaluated using a molecular-mechanics force field. If the new conformation has a lower energy than the initial one, the new conformation is accepted and becomes the starting conformation for the next random dihedral-angle change. If the new conformation has a higher energy, a random number  $r$  that lies between 0 and 1 is generated, and the new conformation is rejected unless  $e^{-|\Delta E|/kT} > r$ , where  $|\Delta E|$  is the energy difference between

the two successive conformations and  $k$  is Boltzmann's constant. This acceptance rule is the **Metropolis criterion**, named after the person who proposed it in conjunction with simulations designed to calculate the thermodynamic properties of fluids. One generates several thousand new conformations using this procedure, which produces a set of conformations whose energies are distributed according to the Boltzmann distribution law. From this set of conformations, one takes the  $n$ th, the  $2n$ th, the  $3n$ th, . . . , where  $n$  is typically 200, to give a few hundred conformations, each of which is then subjected to energy minimization.

Annealing consists of heating a solid to a high temperature, holding it at that temperature for a while, and then very slowly cooling it, thereby relieving strains and reaching the global minimum in the Gibbs free energy. Similarly, very slow cooling of a liquid favors formation of a single, highly ordered, low free-energy crystal. In contrast, rapid cooling of a liquid (quenching) leads to a polycrystalline or amorphous solid that is not the global free-energy minimum. (The minimization procedures of Section 15.10 find the nearest local minimum, rather than the global one, and are somewhat analogous to quenching.) **Simulated annealing** is a calculational procedure that uses a "cooling" process to obtain what is hoped will be the global energy minimum.

One can do simulated annealing with either the Metropolis Monte Carlo or the molecular dynamics search procedures. Using the Metropolis Monte Carlo procedure, one starts with a high-energy conformer that is a local minimum and assigns the molecule an elevated temperature  $T_{\text{high}}$  (say, 1000 K or 1500 K). One does several hundred random dihedral angle changes at  $T_{\text{high}}$ , accepting or rejecting each change using the Metropolis criterion. (The number of Monte Carlo steps needed at each temperature increases with an increase in the number of molecular dihedral angles and increases with an increase in the complexity of the PES.) Then the temperature in the acceptance/rejection inequality is reduced by a small amount; typically, one multiplies it by 0.9. Several hundred Metropolis Monte Carlo steps are taken at the new temperature, which is then reduced by multiplying it by 0.9. The whole process is repeated until  $T$  reaches a very low temperature  $T_{\text{low}}$ , typically 50 K. (Also, when  $T$  is low enough so that a large percentage of the random dihedral changes lead to rejected conformations, one limits the size of the allowed dihedral-angle changes.) If the simulated annealing has been done appropriately for the molecule being studied, the conformation obtained at  $T_{\text{low}}$  will be the global minimum and no energy minimization will be needed. In practice, one often applies an energy-minimization procedure (Section 15.10) to the final  $T_{\text{low}}$  conformation. One then repeats the whole process several times, starting with a new initial conformation at  $T_{\text{high}}$ . If the same final conformation is found on the majority of the runs, one can have some confidence that the global minimum has been found.

The Metropolis Monte Carlo simulated annealing method gave excellent results when applied to a decapeptide containing 10 alanine residues. With 1000 steps taken at each  $T$ , 14 out of 20 runs reached the same minimum, which is therefore believed to be the global minimum [S. R. Wilson and W. Cui, *Biopolymers*, **29**, 225 (1990)]. However, this decapeptide with all its residues identical has a very symmetrical PES and is not a stringent test of simulated annealing. When Metropolis Monte Carlo simulated annealing was applied to the pentapeptide Met-enkephalin, it failed, reaching 24 different conformations on 24 runs [A. Nayeem, J. Vila, and H. A. Scheraga, *J. Comput. Chem.*, **12**, 594 (1991)]. (Met-enkephalin, with 24 single-bond dihedral angles, is a naturally occurring opiate neurotransmitter that has been widely used in conformational searches. Several methods have given the same global minimum for Met-enkephalin.)

Using the molecular dynamics approach for simulated annealing, after the atomic positions have evolved for several hundred time steps, one cools the system a bit by multiplying each velocity component by a scale factor that is slightly less than one. After another several hundred time steps, one again reduces the velocity components. This procedure is repeated until the temperature is very low, typically 50 K.

Simulated annealing is widely used to refine biomolecular structures found from X-ray crystallography or NMR.

In the **diffusion-equation method** (DEM), one starts with a molecular-mechanics expression for the PES  $U$  as a function of the nuclear coordinates. One applies a certain mathematical operator  $\hat{B}$  to  $U$  so as to smooth out its maxima and minima.  $\hat{B}$  is a function of a parameter  $t$  that is analogous to time. As  $t$  increases, more and more minima in  $U$  disappear. One sets the value of  $t$  equal to a time  $t_0$  at which one believes that only one minimum remains. The nature of  $\hat{B}$  is such that the derivative  $\partial U/\partial t$  obeys a differential equation that has the same form as the equation describing the process of diffusion of a solute in solution. One therefore solves this diffusion differential equation to find the smoothed PES  $U_{\text{smoothed}}$  at  $t_0$ . The smoothing process not only removes minima but also changes the locations of the minima that remain. One hopes and prays that  $U_{\text{smoothed}}$  has only one remaining minimum and that this minimum originated from the global minimum in the original PES. With only one minimum in  $U_{\text{smoothed}}$ , this minimum is easily located using one of the methods of Section 15.10. One then reverses the process and mathematically transforms  $U_{\text{smoothed}}$  back to the original  $U$  in several steps, giving the series of functions  $U_{\text{smoothed}}, U_{\text{smoothed}-1}, U_{\text{smoothed}-2}, \dots, U$ . One searches  $U_{\text{smoothed}-1}$  for a minimum, starting the search at the location of the minimum in  $U_{\text{smoothed}}$ . Then the  $U_{\text{smoothed}-2}$  surface is searched starting at the location of the minimum on  $U_{\text{smoothed}-1}$ . And so on. Thus the minimum on the  $U_{\text{smoothed}}$  surface is traced back to where it occurred on the  $U$  surface. The DEM is quite fast. For example, in only 10 minutes of supercomputer time it found a structure close to the global minimum of Met-enkephalin on a certain molecular-mechanics PES [J. Kostrowicki and H. Scheraga, *J. Phys. Chem.*, **96**, 7442 (1992)]. A variation of the DEM procedure is to transform the PES to leave a few minima and then trace each of these back to the original PES. DEM is an example of a **potential-smoothing** (PS) search method. Other PS methods exist.

Unfortunately, the minimum found by the DEM is often not the global minimum, so DEM by itself is an unsatisfactory method. A two-step procedure has been proposed in which one first uses the DEM to locate a minimum and then searches the coordinate space in the region of this minimum. By confining the second search to the region near the DEM minimum, one greatly reduces the ranges of variables that must be searched [S. Nakamura et al., *J. Phys. Chem.*, **99**, 8374 (1995)]. Moreover, one can also do searching in the regions of the minima found during the various steps of the DEM reversal process.

In the **low-mode** conformational search method (LMOD or LMCS), one starts with a minimum-energy conformer and calculates the  $3N - 6$  normal vibrational modes (Section 15.12). Vibrational modes that involve torsion about single bonds have low frequencies, so the method uses only modes having vibrational wavenumbers less than  $250 \text{ cm}^{-1}$ . One moves the atoms from their minimum-energy position along the paths they would follow in one of the low-frequency vibrational modes until a steep energy increase is seen. The high-energy structure that results is then subjected to energy minimization in the hope that a potential-energy barrier will be crossed in the minimization process, thereby leading to a new minimum-energy conformer. This procedure is repeated first for each low-frequency mode of the original conformer, and then is repeated for random mixtures of the low-frequency modes. The low-mode search procedure is then applied to each new local-minimum conformer that has been found. An evaluation of the ability of several molecular-mechanics programs to find the known conformations of protein-bound ligands found the low-mode conformational search method in the MacroModel program (Section 17.5) to perform well [J. Boström, *J. Comput. Aid. Mol. Des.*, **15**, 1137 (2001)].

For any of the search methods used, all conformers found should be verified to be local minima, rather than saddle points, by calculating their vibrational frequencies and checking that these are all real.

The ring hydrocarbon cycloheptadecane ( $C_{17}H_{34}$ ) was used to test several methods of conformational searching. At comparable search times, the following numbers of local-minimum conformers were found that have an energy within 3 kcal/mol of the global minimum on the PES of a certain molecular-mechanics potential-energy function, called the MM2 PES: random dihedral search—249; random Cartesian search—222; systematic search—211; distance geometry—176; molecular dynamics—169 [M. S. Saunders et al., *J. Am. Chem. Soc.*, **112**, 1419 (1990)]. Combining the search results gave a total of 262 minima in this energy range. Although this study found random searching to do better than a systematic search, a subsequent study of  $C_{17}H_{34}$  using an improved systematic search method called SUMM (systematic unbounded multiple minima) found that SUMM was superior to all the methods used in the previous study. SUMM found all 262 minima [I. Kolossváry and W. C. Guida, *J. Comput. Chem.*, **14**, 691 (1993)]. In a comparison of SUMM with LMOD, SUMM took 52 hours of workstation time to find all 262  $C_{17}H_{34}$  low-energy conformers, whereas LMOD accomplished this task in only 28 hours [I. Kolossváry and W. C. Guida, *J. Am. Chem. Soc.*, **118**, 5011 (1996); *J. Comput. Chem.*, **20**, 1671 (1999)].

Another study of  $C_{17}H_{34}$  used a version of DEM plus local searching called PS-NMLS (potential-smoothing with normal-mode local searching) and found (in 13 days of workstation time) a total of 20469  $C_{17}H_{34}$  conformers on the MM2 PES [R.V. Pappu et al., *J. Phys. Chem. B*, **102**, 9725 (1998)]. The energy distribution of these conformers was roughly Gaussian-shaped, with a peak in the 8 to 9 kcal/mol range and maximum conformer energy about 25 kcal/mol above the global minimum.

Most medium-size flexible molecules exist as an equilibrium mixture of a great number of conformers whose shapes can differ considerably from one another. In contrast, a globular protein in its native, biologically active state has a well-defined three-dimensional shape that is determined by intramolecular interactions such as electrostatic attractions between charged amino and carboxylate groups, hydrogen bonding, and dispersion attractions and by interactions with the solvent. (There are significant exceptions to this statement; about 30% of human proteins contain large disordered, unstructured regions; see S. Everts, *Chem. Eng. News*, April 2, 2007, p. 58.) The conformation of a structured protein does fluctuate with time, but the fluctuations are not great and the overall shape is maintained.

A major problem is the protein-folding problem. Given the astronomical number of possible conformations of a protein molecule, how does a protein fold into its native state when it is synthesized, and how can we predict the three-dimensional structure of a protein knowing only its amino acid sequence? The native state of a protein is believed to be the Gibbs free-energy minimum of the protein in solution. If we neglect both the entropy contribution to  $G$  and the effect of the solvent, the native state would be the global energy minimum (GEM). Hence people are deeply interested in finding the GEM of a protein. Methods used to search for the GEM of a protein include simulated annealing, DEM, genetic algorithms, and many others; see D. J. Wales, *Energy Landscapes*, Cambridge Univ. Press, 2003, Sections 6.7 and 9.2.

In addition to the problem of the huge number of possible conformations, inaccuracies in currently used molecular-mechanics force fields (Section 17.5) are also a stumbling block. An analysis of errors involved when calculating protein-folding energies found that as protein size increases, random and systematic errors propagate throughout the protein and distort its potential-energy surface so much that current methods “have little chance of success in finding the free energy minima of large proteins” [J. C. Faver et al., *PLOS ONE*, **6**, e18868 (2011); [www.ncbi.nlm.nih.gov/pmc/articles/PMC3081830](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3081830)].

A 2012 review of protein structure prediction noted that so far “no success” has been achieved in predicting protein structures solely by finding the conformations of lowest free energy, due to the difficulty of accurately modeling intra-protein and protein-solute

interactions at the atomic level and to the huge number of conformations that must be searched; the success that has been achieved is done by using information from known structures of related proteins [A. Roy and Y. Zhang, “Protein Structure Predictions” in *Encyclopedia of Life Sciences*, Wiley, [www.els.net](http://www.els.net) (available at [zhanglab.ccmb.med.umich.edu/papers/2012\\_12.pdf](http://zhanglab.ccmb.med.umich.edu/papers/2012_12.pdf))].

## 15.12 Molecular Vibrational Frequencies

Geometry optimization (Section 15.10) yields a quantum-mechanical estimate of the molecular electronic energy  $U$  evaluated at a local minimum, and a conformational search (Section 15.11) yields an estimate of the global energy minimum. However, the nuclei in a molecule vibrate about their equilibrium positions, and it is essential to include the molecular vibrational zero-point energy  $E_{\text{ZPE}}$  if accurate quantum-mechanical estimates of energy differences are wanted. Calculation of  $E_{\text{ZPE}}$  requires knowing the molecular vibrational frequencies. Also, theoretical calculation of vibrational frequencies helps in analyzing infrared spectra; “it is virtually impossible to interpret and correctly assign the vibrational spectra of larger polyatomic molecules without quantum-mechanical calculations” (P. Pulay in *Yarkony*, Part II, Chapter 19). Finally, calculation of vibrational frequencies allows one to classify a stationary point on the PES found by a geometry-optimization method as a local minimum (all real vibrational frequencies) or an  $n$ th-order saddle-point ( $n$  imaginary frequencies).

The Schrödinger equation for nuclear motion in a molecule is given by Eqs. (13.10) and (13.11) as  $\hat{H}_N\psi_N = (\hat{T}_N + U)\psi_N = E\psi_N$ . The nuclear Schrödinger equation for diatomic molecules was solved approximately in Section 13.2. For polyatomic molecules, derivations will be omitted. (For discussion of polyatomic-molecule vibrations, see *Wilson, Decius, and Cross*.) The total molecular energy  $E$  is approximately the sum of translational, rotational, vibrational, and electronic energies. In the harmonic-oscillator approximation, the vibrational energy of an  $N$ -atom molecule is the sum of  $3N - 6$  normal-mode vibrational energies ( $3N - 5$  for a linear molecule):

$$E_{\text{vib}} \approx \sum_{k=1}^{3N-6} (v_k + \frac{1}{2})h\nu_k \quad (15.74)$$

where  $\nu_k$  is the **harmonic (or equilibrium) vibrational frequency** for the  $k$ th normal mode and each vibrational quantum number  $v_k$  has the possible values 0, 1, 2, . . . independent of the values of the other vibrational quantum numbers. For the ground vibrational state, each of the  $3N - 6$  vibrational quantum numbers equals zero, and the zero-point energy in the harmonic-oscillator approximation is  $E_{\text{ZPE}} \approx \frac{1}{2} \sum_{k=1}^{3N-6} h\nu_k$ .

The harmonic vibrational frequencies of a molecule are calculated as follows. (1) Solve the electronic Schrödinger equation  $(\hat{H}_{\text{el}} + V_{\text{NN}})\psi_{\text{el}} = U\psi_{\text{el}}$  for several molecular geometries to find the equilibrium geometry of the molecule (Section 15.10). (2) Calculate the set of second derivatives  $(\partial^2 U / \partial X_i \partial X_j)_e$  of the molecular electronic energy  $U$  with respect to the  $3N$  nuclear Cartesian coordinates of a coordinate system with origin at the center of mass, where these derivatives (the Hessian matrix elements—Section 15.10) are evaluated at the equilibrium geometry. These second derivatives can be evaluated analytically from ab initio SCF MO wave functions (and many other wave functions), although their ab initio calculation is time-consuming. (3) Form the **mass-weighted force-constant** (or mass-weighted Hessian) **matrix elements**

$$F_{ij} \equiv \frac{1}{(m_i m_j)^{1/2}} \left( \frac{\partial^2 U}{\partial X_i \partial X_j} \right)_e \quad (15.75)$$

where  $i$  and  $j$  each go from 1 to  $3N$  and  $m_i$  is the mass of the atom corresponding to coordinate  $X_i$ . (4) Solve the following set of  $3N$  linear equations in  $3N$  unknowns:

$$\sum_{j=1}^{3N} (F_{ij} - \delta_{ij}\lambda_k)l_{jk} = 0, \quad i = 1, 2, \dots, 3N \quad (15.76)$$

In this set of equations,  $\delta_{ij}$  is the Kronecker delta, and  $\lambda_k$  and the  $l_{jk}$ 's are as-yet unknown parameters whose significance will be seen shortly. In order that this set of homogeneous equations have a nontrivial solution, the coefficient determinant must vanish:

$$\det(F_{ij} - \delta_{ij}\lambda_k) = 0 \quad (15.77)$$

This determinant is of order  $3N$  and when expanded gives a polynomial whose highest power of  $\lambda_k$  is  $\lambda_k^{3N}$ , so the determinantal (secular) equation will yield  $3N$  roots (some of which may be the same) for  $\lambda_k$ . The molecular harmonic vibrational frequencies are then calculated from

$$\nu_k = \lambda_k^{1/2}/2\pi$$

Six of the  $\lambda_k$  values found by solving (15.77) will be zero, yielding six frequencies with value zero, corresponding to the three translational and three rotational degrees of freedom of the molecule. (In practice, because the equilibrium geometry is never found with infinite accuracy, one may find six vibrational frequencies with values close to zero:  $|\nu_k|/c < 30 \text{ cm}^{-1}$ .) The remaining  $3N - 6$  vibrational frequencies are the molecular harmonic vibrational frequencies.

Note that a vibrational-frequency calculation must be preceded by a geometry optimization using the same method and basis set as used for the frequency calculation. "Frequencies" calculated at a point that is not a stationary point are not true vibrational frequencies.

Once the  $\lambda_k$ 's have been found, we solve the set of equations (15.76)  $3N - 6$  times, each time with a different one of the nonzero  $\lambda_k$  values, to yield the numbers  $l_{jk}$ . The quantity  $m_s^{1/2}l_{jk}/m_j^{1/2}l_{sk}$  gives the ratio of the classical-mechanical vibrational amplitude of the coordinate  $X_j$  to the amplitude of  $X_s$  for the  $k$ th normal mode. For example, for the diatomic molecule  $^1\text{H}^{19}\text{F}$ , there is only one normal mode ( $k = 1$ ). Letting the coordinates  $X_1, \dots, X_6$  be  $x_{\text{H}}, y_{\text{H}}, z_{\text{H}}, x_{\text{F}}, y_{\text{F}}, z_{\text{F}}$ , respectively, with origin at the center of mass and  $z$  axis through the nuclei, one finds on solving (15.76) (Levine, *Molecular Spectroscopy*, Section 6.2) that  $l_{31}/l_{61} = -(m_{\text{F}}/m_{\text{H}})^{1/2}$ . Hence  $m_6^{1/2}l_{31}/m_3^{1/2}l_{61} = -m_{\text{F}}/m_{\text{H}} = -19$ . Thus the vibrational amplitude (maximum displacement from equilibrium) of the H nucleus in HF is 19 times the amplitude of the F nucleus, with the vibrational displacements being in opposite directions. Figure 15.16 shows the forms of the normal modes of the water molecule.

In matrix notation, the equations (15.76) are  $\mathbf{F}\mathbf{L}^{(k)} = \lambda_k\mathbf{L}^{(k)}$ , where  $\mathbf{F}$  has matrix elements  $F_{ij}$  and  $\mathbf{L}^{(k)}$  is a column vector with elements  $l_{jk}$ . Thus the  $\lambda_k$ 's are eigenvalues of the mass-weighted force-constant matrix, and are found by the usual matrix methods (Section 8.6).

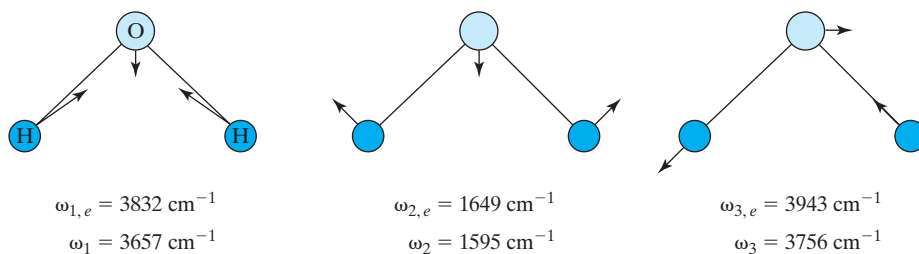


**FIGURE 15.16** Normal modes of  $\text{H}_2\text{O}$ . The harmonic ( $\omega_e$ ) and fundamental ( $\omega$ ) wavenumbers of the modes are given.

The observed light-absorption frequency for the transition in which the vibrational quantum number  $v_k$  goes from 0 to 1 with no change in the other vibrational quantum numbers is called the **fundamental** (or **anharmonic**) **frequency** for the  $k$ th normal mode. Because of vibrational anharmonicity, a fundamental frequency is smaller than the corresponding harmonic frequency [recall Eq. (4.62) for diatomic molecules]. Vibrational frequencies are converted to wavenumbers by being divided by  $c$  [ $\tilde{\nu} \equiv 1/\lambda = \nu/c$ , Eq. (4.64)] and wavenumbers are usually expressed in the units  $\text{cm}^{-1}$ . (Chemists have the bad habit of referring to the  $\tilde{\nu}$  values as “frequencies.”) Fundamental vibrational wavenumbers are usually symbolized by  $\omega$ . The harmonic and fundamental wavenumbers of gas-phase  $\text{H}_2\text{O}$  are shown in Fig. 15.16. Experimental values of harmonic vibrational frequencies are calculated from observed fundamental frequencies using anharmonicity constants found from analysis of vibrational infrared and Raman spectra. For medium and large molecules, often only the fundamental frequencies are known. For calculating  $E_{\text{ZPE}}$  from  $\frac{1}{2} \sum_k h\nu_k$ , more accurate results are obtained by using the fundamental frequencies rather than the harmonic frequencies, since the fundamental frequencies incorporate anharmonicity corrections.

Harmonic vibrational frequencies calculated by the Hartree–Fock method are usually several percent higher than observed harmonic frequencies. Pretty good estimates of experimental *fundamental* frequencies can be found by multiplying ab initio SCF MO *harmonic* frequencies by an empirically found **scale factor**. For HF/6-31G\* calculations, the vibrational scale factor is 0.895 [A. P. Scott and L. Radom, *J. Phys. Chem.*, **100**, 16502 (1996); this paper gives scale factors for many methods and basis sets, including separate scale factors for vibrational frequencies, zero-point energies, and thermodynamic properties]. The Scott–Radom study found that with the 0.895 scale factor, 84% of HF/6-31G\* fundamental frequencies of a sample of 122 molecules were within 6% of the experimental value and recommended this basis set for HF vibrational-frequency calculations. For example, HF/6-31G\* harmonic wavenumbers for  $\text{H}_2\text{O}$  are 4070, 1826, and  $4188 \text{ cm}^{-1}$ . With the 0.895 scale factor, the predicted HF/6-31G\* fundamental wavenumbers of gas-phase  $\text{H}_2\text{O}$  are 3643, 1634, and  $3748 \text{ cm}^{-1}$ , in good agreement with the fundamentals in Fig. 15.16. The CCCBDB ([cccbdb.nist.gov](http://cccbdb.nist.gov)) lists vibrational scale factors for a variety of methods and basis sets; see also J. M. Alecu et al., *J. Chem. Theory Comput.*, **6**, 2872 (2010) (available at [static.msi.umn.edu/rreports/2010/163.pdf](http://static.msi.umn.edu/rreports/2010/163.pdf)).

## 15.13 Thermodynamic Properties

This section discusses how gas-phase thermodynamic properties can be calculated from molecular quantities such as the electronic energy, the equilibrium geometry, and the vibrational frequencies.

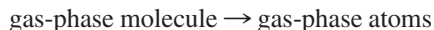
The (total) ground-state dissociation energy  $D_0$  of a molecule is the energy needed to dissociate the molecule in its ground vibrational state to atoms in their ground states.  $D_0$  differs from the equilibrium dissociation energy  $D_e$  by the zero-point vibrational energy  $E_{\text{ZPE}}$ :

$$D_0 = D_e - E_{\text{ZPE}} \approx D_e - \frac{1}{2} h \sum_{k=1}^{3N-6} \nu_k \quad (15.78)$$

where fundamental, rather than harmonic, vibrational frequencies are used in  $E_{\text{ZPE}}$ . To calculate  $D_e$ , one calculates the molecular electronic energy  $U_e$  at the equilibrium geometry, calculates the ground-state energy of each atom in the molecule using the same method and basis set used in the molecular calculation, and takes the difference between the total atomic energies and the molecular energy. Even though the zero-point energy of a single vibrational mode is rather small, a medium-size or large molecule has many vibrational

modes and its  $E_{\text{ZPE}}$  is substantial. For example, 1,3-butadiene,  $\text{CH}_2\text{CHCHCH}_2$ , with 24 normal modes, has  $E_{\text{ZPE}} = 2.2$  eV, which corresponds to 50 kcal/mol.

Recall from Section 13.1 that  $N_{\text{A}}D_0 = \Delta U_0^\circ = \Delta H_0^\circ$  for the process



This process is called *atomization* (at), and the 0 K gas-phase **atomization energy** is  $\Delta U_{\text{at},0}^\circ = N_{\text{A}}D_0$ .

For example, for  $\text{H}_2\text{O}$ , a geometry-optimized HF/6-31G\* calculation gives  $U_e = -76.010746$  hartrees  $= -76.010746E_{\text{h}}$ . In doing energy-difference calculations involving nonsinglet species (such as most ground-state atoms), one usually uses the UHF rather than the ROHF energy (Section 15.3). UHF/6-31G\* ground-state atomic energies are  $-0.498233E_{\text{h}}$  for H and  $-74.783931E_{\text{h}}$  for O. The predicted  $D_e$  is then  $D_e/E_{\text{h}} = 2(-0.498233) + (-74.783931) - (-76.010746) = 0.23035$  and Table A.2 gives  $D_e = 6.27$  eV. The HF/6-31G\* scaled fundamental vibrational wavenumbers of  $\text{H}_2\text{O}$  are (Section 15.12) 3643, 1634, and  $3748\text{ cm}^{-1}$ , which give  $E_{\text{ZPE}} \approx 0.56$  eV and  $D_0 = 5.71$  eV, as compared with the experimental value  $D_0 = 9.51$  eV. The  $D_0 = 5.71$  eV value when multiplied by the Avogadro constant  $N_{\text{A}}$  gives a predicted atomization energy  $\Delta U_{\text{at},0}^\circ = 551\text{ kJ/mol} = 132\text{ kcal/mol}$ . To compare this predicted value with experiment, one consults a table of thermodynamic data. Gas-phase enthalpies of formation  $\Delta H_{f,0}^\circ = \Delta U_{f,0}^\circ$  at 0 K are 216.04 kJ/mol for H(g), 246.79 kJ/mol for O(g), and  $-238.92\text{ kJ/mol}$  for  $\text{H}_2\text{O(g)}$  (Chase *et al.*). These data give the experimental atomization energy as  $\Delta U_{\text{at},0}^\circ = 917.79\text{ kJ/mol} = 219.4\text{ kcal/mol}$ . Accurate calculation of dissociation energies requires inclusion of electron correlation (see Chapter 16).

Enthalpies of formation are more commonly used than atomization energies and are readily calculated from a predicted atomization energy and known thermodynamic data for the elements. For  $\text{H}_2\text{O(g)}$ ,  $\Delta H_f^\circ$  is  $\Delta H^\circ$  for the reaction  $\text{H}_2(\text{g}) + \frac{1}{2}\text{O}_2(\text{g}) \rightarrow \text{H}_2\text{O(g)}$ . Using the scheme reactants  $\rightarrow$  atoms  $\rightarrow$  products, we can calculate  $\Delta H^\circ$  for the formation reaction by subtracting the theoretically predicted atomization energy of the product from the experimental atomization energies of the reactants. Using the data in the preceding paragraph, we have as the HF/6-31G\* prediction:  $\Delta H_{f,0}^\circ(\text{H}_2\text{O(g)}) = [2(216.04) + 246.79 - 551]\text{ kJ/mol} = 128\text{ kJ/mol} = 30.6\text{ kcal/mol}$ . The experimental value is  $-238.9\text{ kJ/mol} = -57\text{ kcal/mol}$ . The error is, of course, the same as the error in the atomization energy.

So far, we have worked at 0 K. Values at 298 K are of more interest. Statistical mechanics (see any physical chemistry text) gives the translational contribution to the molar internal energy of an ideal gas as  $\frac{3}{2}RT$ , the rotational contribution as  $RT$  for linear molecules and  $\frac{3}{2}RT$  for nonlinear molecules, and the vibrational contribution by a formula given in Prob. 15.39. (The electronic contribution is negligible at room temperature except for a few cases of molecules with low-lying excited electronic states.) One finds that only vibrational modes with wavenumbers below  $900\text{ cm}^{-1}$  contribute significantly to the room-temperature vibrational internal energy (Prob. 15.39). For  $\text{H}_2$ ,  $\text{O}_2$ , and  $\text{H}_2\text{O}$ , which have no low-frequency vibrations, the vibrational contributions to the internal energy are negligible at room temperature. For the formation reaction  $\text{H}_2(\text{g}) + \frac{1}{2}\text{O}_2(\text{g}) \rightarrow \text{H}_2\text{O(g)}$ , the molar internal energy of the product is  $\frac{3}{2}RT + \frac{3}{2}RT = 3RT$  higher at  $T$  than at 0 K, and the molar internal energy of the reactants is  $\frac{3}{2}RT + RT + \frac{1}{2} \cdot \frac{3}{2}RT + \frac{1}{2}RT = 3.75RT$  higher at  $T$  than at 0. In addition, the formation reaction has a net change of  $-\frac{1}{2}$  mole of gas, and the relations  $H \equiv U + PV = U + nRT$  tell us that

$$\Delta H_{f,T}^\circ = \Delta U_{f,T}^\circ - \frac{1}{2}RT = \Delta U_{f,0}^\circ - 0.75RT - \frac{1}{2}RT = \Delta U_{f,0}^\circ - \frac{5}{4}RT = \Delta H_{f,0}^\circ - \frac{5}{4}RT$$

Thus the HF/6-31G\* predicted value 30.6 kcal/mol for  $\Delta H_{f,0}^\circ$  of  $\text{H}_2\text{O(g)}$  corresponds to a predicted  $\Delta H_{f,298}^\circ$  of 29.9 kcal/mol.

Statistical mechanics gives the molar entropy of an ideal gas as the sum of translational, rotational, vibrational, and electronic contributions. (See any physical chemistry text.) The translational contribution depends only on the molar mass of the gas. The rotational contribution  $S_{\text{rot}}$  depends on the symmetry number and the principal moments of inertia. These quantities are readily found from the molecule's equilibrium geometry. The vibrational contribution  $S_{\text{vib}}$  depends on the molecular vibrational frequencies, which can usually be rather accurately calculated with the aid of a scale factor. The electronic contribution depends on the electronic degeneracy of the ground electronic state and in a few cases on the energies of any low-lying excited electronic states. Thus, ab initio quantum chemistry calculations can give accurate predictions of gas-phase room-temperature entropies for small molecules.

For example, East and Radom devised a procedure they call E1, which calculates  $S_{\text{rot}}$  from the MP2/6-31G\* geometry (MP2 calculations are discussed in Section 16.3) and  $S_{\text{vib}}$  from HF/6-31G\* scaled vibrational frequencies and the harmonic-oscillator approximation, except that internal rotations with barriers less than  $1.4RT$  are treated as free rotations [A. L. L. East and L. Radom, *J. Chem. Phys.*, **106**, 6655 (1997)]. For 19 small molecules with no internal rotors, their E1 procedure gave gas-phase  $S_{\text{m},298}^{\circ}$  values with a mean absolute deviation from experiment of only 0.2 J/mol-K and a maximum deviation of 0.6 J/mol-K. The E1 procedure was in error by up to  $1\frac{1}{2}$  J/mol-K for molecules with one internal rotor and by up to 2 J/mol-K for molecules with two rotors. An improved procedure called E2 replaces the harmonic-oscillator potential for internal rotors by a cosine potential calculated using the MP2 method and a large basis set, and reduces the error to 1 J/mol-K for one-rotor molecules.

Improved accuracy in thermodynamic quantities is obtained if anharmonicity corrections (calculated using density functional theory) are included and if internal rotation about single bonds is properly treated; see M. Barone, *J. Chem. Phys.*, **120**, 3059 (2004), and this procedure is included in the *Gaussian* program (Section 15.14).

One is often interested in energy differences between two species B and C, such as different isomers or different conformers. The quantity  $N_{\text{A}}(U_{\text{e},\text{B}} - U_{\text{e},\text{C}})$  gives the molar internal-energy difference at 0 K with zero-point energies neglected. The quantity  $N_{\text{A}}[(U_{\text{e},\text{B}} + E_{\text{ZPE},\text{B}}) - (U_{\text{e},\text{C}} + E_{\text{ZPE},\text{C}})]$  gives the molar internal-energy difference at 0 K. Different conformers usually have similar zero-point vibrational energies, so the zero-point-energy contribution is often neglected here. Translational, rotational, and vibrational contributions are included to get energy differences at temperatures warmer than 0 K. As an example, a high-level ab initio calculation that included electron correlation found the following energy (and enthalpy) differences between the gauche and anti conformers of butane(g): 0.59 kcal/mol at 0 K with zero-point energy neglected, 0.70 kcal/mol at 0 K with zero-point energy included, and 0.64 kcal/mol at 298 K, as compared with an experimental value 0.67 kcal/mol averaged over the range 220 to 298 K [G. D. Smith and R. L. Jaffe, *J. Phys. Chem.*, **100**, 18718 (1996)].

## 15.14 Ab Initio Quantum Chemistry Programs

This section surveys some of the available ab initio quantum chemistry program packages.

The program *Gaussian* ([www.gaussian.com](http://www.gaussian.com)), which exists in various versions (... *Gaussian 94*, *Gaussian 98*, *Gaussian 03*, *Gaussian 09*,...) labeled by release year, is a widely used versatile program package that includes all common ab initio methods, such as Hartree–Fock, CI, MCSCF, MP (Section 16.3), and CC (Section 16.4), the density functional method, many semiempirical methods, and the molecular-mechanics method. *Gaussian* can optimize geometries, calculate vibrational frequencies, thermodynamic properties, and NMR shielding constants, search for transition states, calculate MEPs, and include the effects of a solvent. *Gaussian* is available in versions for supercomputers, workstations, PCs running Windows, and Macintosh computers.

GAMESS (General Atomic and Molecular Electronic Structure System) includes all the common ab initio, density functional, and semiempirical methods and has the advantage of being free. It runs on supercomputers, workstations, Macintoshes, and Windows PCs. For details, see [www.msg.chem.iastate.edu/gamess](http://www.msg.chem.iastate.edu/gamess); M. W. Schmidt et al., *J. Comput. Chem.*, **14**, 1347 (1993); M. S. Gordon and M. W. Schmidt, “Advances in electronic structure theory: GAMESS a decade later,” in *Theory and Applications of Computational Chemistry*, C. E. Dykstra et al. eds., Elsevier 2005. GAMESS is sometimes called GAMESS-US to distinguish it from a different program GAMESS-UK.

Q-Chem ([www.q-chem.com](http://www.q-chem.com)) is an ab initio package for computers using UNIX or LINUX and can also be run on Windows and Macintosh machines by using the Spartan program (see below) as a front end. Q-Chem allows calculations on large molecules and can do Hartree–Fock, MP, CC, and density-functional calculations [Y. Shao et al., *Phys. Chem. Chem. Phys.*, **8**, 3172 (2006)]. It incorporates methods such as CFMM and ONX to achieve linear scaling (Section 15.16) for large molecules.

Jaguar ([www.schrodinger.com](http://www.schrodinger.com)) runs on Linux, Windows, or Macintosh and can do HF, MP2, density functional, and GVB (Section 16.13) calculations.

NWChem ([www.nwchem-sw.org](http://www.nwchem-sw.org)) is a free ab initio and density functional program designed to run on parallel supercomputers and workstation clusters and also runs on Windows and Macintosh computers [M. Valiev et al., *Comput. Phys. Commun.*, **181**, 1477 (2010)].

The ACES III program ([www.qtp.ufl.edu/ACES](http://www.qtp.ufl.edu/ACES)) is a free program designed for performing CC and MP (Sections 16.3 and 16.4) calculations on parallel machines.

Turbomole ([www.cosmologic.de/QuantumChemistry/main\\_qChemistry.html](http://www.cosmologic.de/QuantumChemistry/main_qChemistry.html)) is an ab initio and density functional program that runs on workstations and Windows and Macintosh computers.

Molpro ([www.molpro.net](http://www.molpro.net)) is an ab initio and density functional program designed for highly accurate calculations on small- and medium-size molecules and includes many electron-correlation methods.

ORCA ([www.mpibac.mpg.de/bac/logins/neese/description.php](http://www.mpibac.mpg.de/bac/logins/neese/description.php)) is a UNIX and Windows ab initio, density functional, and semiempirical program that is free to academic users [F. Neese, *WIREs Comput. Mol. Sci.*, **2**, 73 (2012)].

PSI4 ([www.psicode.org](http://www.psicode.org)) is a free ab initio and density functional program that runs on Linux and Macintosh machines [J. M. Turney et al., *WIREs Comput. Mol. Sci.*, **2**, 556 (2012)].

Spartan ([www.wavefun.com](http://www.wavefun.com)) includes ab initio (Hartree–Fock, MP, CC, CI), density functional, semiempirical, and molecular-mechanics methods, has several conformational searching methods, and runs on workstations, PCs, and Macintoshes. Spartan Student Edition is an inexpensive version with fewer capabilities.

HyperChem ([www.hyper.com](http://www.hyper.com)) includes Hartree–Fock, MP2, density-functional, semiempirical, and molecular-mechanics methods and runs on PCs, Macs, and workstations. Student HyperChem is an inexpensive student version that limits the size of the molecule that can be calculated.

Large-scale quantum chemistry calculations can be economically done on a cluster of networked off-the-shelf PCs running in parallel. Such clusters (which typically contain from 8 to 128 PCs) are either put together by the researcher or can be bought.

## 15.15 Performing Ab Initio Calculations

This section discusses some practical matters in using ab initio calculations.

### Input

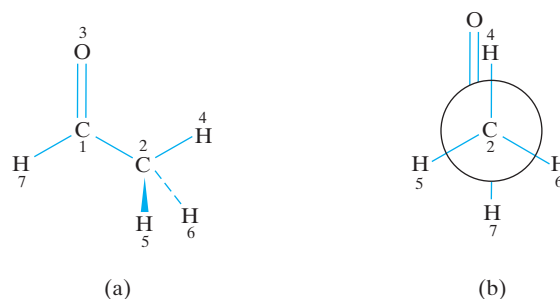
The input section of a calculation specifies the molecule, the calculational method to be used, the basis set, the kind of calculation (single-point, geometry optimization, frequency, and so on), and a molecular geometry (which will be used for a single-point calculation

**TABLE 15.7** Input for *Gaussian* CH<sub>3</sub>CHO  
Geometry Optimization

|                                                    |   |      |   |       |   |        |
|----------------------------------------------------|---|------|---|-------|---|--------|
| # RHF/3-21G Opt                                    |   |      |   |       |   |        |
|                                                    |   |      |   |       |   |        |
| Acetaldehyde (H eclipsing O) HF/3-21G optimization |   |      |   |       |   |        |
|                                                    |   |      |   |       |   |        |
| 0                                                  | 1 |      |   |       |   |        |
| C1                                                 |   |      |   |       |   |        |
| C2                                                 | 1 | 1.52 |   |       |   |        |
| O3                                                 | 1 | 1.22 | 2 | 120.0 |   |        |
| H4                                                 | 2 | 1.09 | 1 | 109.5 | 3 | 0.0    |
| H5                                                 | 2 | 1.09 | 1 | 109.5 | 3 | 120.0  |
| H6                                                 | 2 | 1.09 | 1 | 109.5 | 3 | −120.0 |
| H7                                                 | 1 | 1.08 | 2 | 120.0 | 3 | 180.0  |

or optimized in an optimization calculation). For example, **Table 15.7** shows the *Gaussian* input section for a geometry optimization of the H-eclipsing-O conformer of acetaldehyde, CH<sub>3</sub>CHO (**Fig. 15.17**), using the Hartree–Fock method with the 3-21G basis set. The first line specifies the restricted Hartree–Fock method (the letters HF could have been used instead of RHF) and the 3-21G basis set. The keyword Opt requests a geometry optimization. If no keyword were present, a single-point calculation would be done. The second line is blank. The third line is a description of the calculation for the user’s information and does not affect the computation. The fourth line is blank. The first number in the fifth line specifies the molecular charge and the second number gives the spin multiplicity  $2S + 1$ . The next seven lines specify the initial guess for the geometry. The last line of the input file is blank. (In the PC version of *Gaussian*, the first three nonblank lines of Table 15.7 are entered into lines labeled Route Section, Title Section, and Charge and Multipl., respectively, and required blank lines are automatically inserted by the program.)

The molecular-geometry guess in Table 15.7 is specified in internal coordinates (bond distances, bond angles, and dihedral angles) in a format called a **Z-matrix**. Each row in a Z-matrix specifies the location of an atom relative to previously specified atoms. The first column of the Z-matrix lists the atoms in the molecule. The numbers 1 through 7 are optional and were included for convenience. The order in which the atoms are listed in the first column is chosen by the user. In any given row (except the first), the third column of the Z-matrix specifies the distance (bond length) in angstroms between the atom in the first column and the atom whose row number the user listed in the second column. For

**FIGURE 15.17** The H-eclipsing-O conformation of CH<sub>3</sub>CHO.

example, the 1.52 in column 3 of row 2 of the Z-matrix specifies the C2 to C1 distance as 1.52 Å and the 1.09 in column 3 of row 5 specifies the H5 to C2 distance as 1.09 Å. (The bond distances and angles were chosen as standard values—see Section 15.10 and Table 15.5.) In any given row of the Z-matrix (except the first and second), the fifth column specifies the angle in degrees for the bond formed by the atoms referred to in columns 1, 2, and 4 of that row, with the column-2 atom at the vertex of the angle. For example, the 120.0 in column 5 of row 3 specifies the angle O3C1C2, and the 109.5 in column 5 of row 4 specifies the angle H4C2C1. (*Gaussian* requires that all bond distances, angles, and dihedral angles have a decimal point.)

In any given row (except the first, second, and third), the entry in column 7 gives the dihedral angle for the four atoms listed in columns 1, 2, 4, and 6, in that order. For example, the 0.0 in column 7 of row 4 is the dihedral angle D(H4, C2, C1, O3). To determine this angle, we draw a Newman projection with the middle two atoms C2 and C1 perpendicular to the plane of the paper (Fig. 15.17). This figure shows that H4 and O3 eclipse each other, which is a 0° dihedral angle. A 120° clockwise rotation is needed to turn the front atom H5 to reach O3, so row 5 has D(H5, C2, C1, O3) equal to 120°. Since a 120° counterclockwise rotation is needed to turn H6 to reach O3, row 6 has D(H6, C2, C1, O3) equal to -120°. Row 7 has D(H7, C1, C2, O3) = 180.0°. This is an unusual dihedral angle in that the end atoms H7 and O3 are both bonded to the same atom, C1. This situation could have been avoided by putting a 4 instead of a 3 in column 6 of row 7, but in a molecule such as H<sub>2</sub>C=O, one cannot avoid a dihedral angle with the end atoms both bonded to the same atom.

A Z-matrix is constructed one row at a time from the assumed geometry. Row 1 contains only a single atom. Row 2 contains a bond length but no bond angle or dihedral angle. Row 3 contains a bond length and a bond angle, but no dihedral angle. Row 4 and all subsequent rows each contain a bond length, a bond angle, and a dihedral angle. The atom numbers used in a given row to specify which bond length, bond angle, or dihedral angle is being defined must all refer to atoms previously specified. For example, in row 4 of the Table 15.7 Z-matrix, we cannot put a 5 in column 4, since atom 5 has not yet been specified. It takes some practice to become good at constructing Z-matrices.

The hardest problem is getting the dihedral angles right. If row *r* of the Z-matrix contains the atoms *r*, *i*, *k*, *d*, in that order, then the last entry in row *r* is the dihedral angle D(*r*, *i*, *k*, *d*), whose value is found as follows: draw a Newman projection with the *i* to *k* bond perpendicular to the plane of the paper and with atom *i* in front of *k*. The dihedral angle D(*r*, *i*, *k*, *d*) is the angle needed to rotate the bond to *r* clockwise until it coincides with the bond to *d*. Chemists usually follow the convention that dihedral angles lie in the range -180° to 180°, so if a counterclockwise rotation is needed to make the bond to *r* coincide with the bond to *d*, then D(*r*, *i*, *k*, *d*) is in the range 0° to -180°.

In specifying a bond angle  $\theta$  in a Z-matrix, the angle must be in the range  $0 < \theta < 180^\circ$ . For a linear molecule such as CO<sub>2</sub>, the forbidden angle of 180° is avoided as follows: One adds a dummy atom, symbolized by X. A convenient location for X is 1.0 Å from C with the fictitious C—X bond perpendicular to the molecular axis. The X atom is ignored in the calculation and simply serves as a reference point in the Z-matrix to define bond angles. When dealing with rings, it's a good idea to put a dummy atom in the center of the ring.

Sometimes one wants to hold one or more geometrical parameters constant during an optimization. For example, to find the electronic energy *U* as a function of the torsional angle in butane (Fig. 15.13), we would do several calculations, in each of which the dihedral angle D(CCCC) is held constant at a different value, while all other internal coordinates are optimized. A procedure for doing this in *Gaussian* is described in Prob. 15.53.

Instead of using bond lengths, angles, and dihedral angles in the Z-matrix, one can specify the nuclear positions using Cartesian coordinates for each atom. For example, for

CO<sub>2</sub>, if we take the *z* axis as the molecular axis and guess a bond length of 1.22 Å, the geometry specification is

```
C   0.0   0.0   0.0
O   0.0   0.0   1.22
O   0.0   0.0  -1.22
```

To obtain Cartesian coordinates for complicated molecules, one can use a molecule builder, as discussed in the next paragraph and in the subsection at the end of this section.

Instead of using a Z-matrix or Cartesian coordinates to input the guessed geometry (which becomes tedious and prone to error for large molecules), many programs have a **molecule builder**, which allows one to construct a ball-and-stick (or wireframe or space-filling) model of the molecule on screen. The model is built from fragments selected by the user. Possible fragments are atoms, groups, rings, and so on. The builder uses standard bond lengths and bond angles and allows one to adjust dihedral angles to get a desired conformer, or the dihedral angle chosen by the builder can be used. One can interrogate the model to get the bond lengths, bond angles, and dihedral angles, and can obtain the Cartesian coordinates of the atoms in the model. Once the model has been constructed, one chooses the desired kind of calculation from a menu. Most molecule builders allow you to use the molecular-mechanics method to improve the structure of the molecule before running the quantum-mechanical calculation. Molecular mechanics (Section 17.5) is a very fast nonquantum-mechanical method.

Several programs exist that have molecule builders that can provide input to programs like *Gaussian* and GAMESS that do not contain graphical interfaces; many of these builders can also visualize results (such as displaying MO shapes and isodensity surfaces, animating normal vibrational modes, displaying predicted spectra, etc.) from the output of *Gaussian* and GAMESS. The program GaussView ([www.gaussian.com](http://www.gaussian.com); Linux, Windows, Macintosh) does this for *Gaussian*, and the Windows program ChemBio3D Ultra ([www.cambridgesoft.com](http://www.cambridgesoft.com)) does this for *Gaussian*, GAMESS (which is included as part of ChemBio3D Ultra), and Jaguar. The program Maestro ([www.schrodinger.com](http://www.schrodinger.com); free to academic users; Linux, Windows, Macintosh) provides a graphical interface for Jaguar. The Windows program Molden ([www.cmbi.ru.nl/molten](http://www.cmbi.ru.nl/molten)), free to academic users, provides a graphical interface for *Gaussian* and GAMESS output. The free program MacMolPlt ([www.scl.ameslab.gov/~brett/MacMolPlt](http://www.scl.ameslab.gov/~brett/MacMolPlt); Linux, Windows, Macintosh) provides a graphical interface for GAMESS. The free program Avogadro ([avogadro.openmolecules.net](http://avogadro.openmolecules.net); Windows, Macintosh, Linux) provides a graphical interface for input and output to a variety of programs (*Gaussian*, GAMESS, Q-Chem, NWChem, MolPro, MOPAC) [M. D. Hanwell et al., *J. Cheminform.*, **4**, 17 (2012)]. The free program Gabedit ([gabedit.sourceforge.net](http://gabedit.sourceforge.net); Windows, Macintosh, Linux) provides graphical input and output for nine quantum chemistry programs [A.-R. Allouche, *J. Comput. Chem.*, **32**, 174 (2011)]. Q-Chem can use the free program IQmol ([iqmol.org](http://iqmol.org)) for graphical input and output. The free program Molekel ([molekel.cscs.ch](http://molekel.cscs.ch); Linux, Windows, Macintosh) will visualize output from *Gaussian* and GAMESS. The free program Ecce ([ecce.emsl.pnl.gov](http://ecce.emsl.pnl.gov)) provides a graphical interface to NWChem and *Gaussian*. The free Windows program Facio ([www1.bbiq.jp/zzzfelis/Facio.html](http://www1.bbiq.jp/zzzfelis/Facio.html)) is a graphical interface to GAMESS and *Gaussian*.

You can use a free online demonstration version of the program WebMO to build a molecule, view the corresponding Z-matrix, and run a variety of computations (provided each computation does not exceed the time limit). Go to [www.webmo.net/demo](http://www.webmo.net/demo) and follow the directions given in the tutorial at the bottom of the page. In addition to providing the energies and structures predicted by several methods, the WebMO demo can calculate vibrational frequencies and show animations of normal vibrational modes, and can display the predicted NMR spectrum, the shapes of MOs, and the molecular surface electrostatic potential.

After a geometry optimization calculation is finished, programs such as WebMO and Spartan Student Edition display a model of the molecule; the calculated bond distances, bond angles, and dihedral angles can be found by clicking on atoms of the model. A common student error in finding bond angles is to click on the atoms in the wrong order. For example, the bond angle in water must be found by clicking in the order HOH.

### Kinds of Calculations

So far in this chapter, single-point, geometry optimization, and frequency calculations have been mentioned (Sections 15.10 and 15.12). Other important kinds of calculations are discussed in later sections (transition-state searches, reaction-path following, PES scans in Section 16.14; inclusion of solvation effects in Section 15.17).

Although it is desirable that molecular geometry be optimized before molecular properties are calculated, ab initio geometry optimizations are impractical for very large molecules with little symmetry. For very large molecules we must be content with single-point ab initio calculations done at either an experimentally determined geometry or at a geometry optimized using a semiempirical or molecular-mechanics method. Geometry optimization with many of the methods that include electron correlation is limited to smaller molecules than can be optimized using the Hartree–Fock (HF) or density-functional method. Since the HF and density-functional methods yield generally reliable geometries, a frequent procedure is to optimize the geometry with the HF or density functional method and then use a correlation method to do a single-point calculation of the molecular energy at the HF or density functional geometry.

Geometry optimization was discussed in Section 15.10, where it was noted that many of the optimization procedures used in quantum chemistry programs can converge to a stationary point that is a saddle point rather than a minimum. To verify that a true minimum has been found, one can follow the geometry optimization with a frequency calculation. All  $3N - 6$  calculated vibrational frequencies must be real to have a minimum. One way to avoid getting a saddle point instead of a minimum is to eliminate all symmetry in the starting geometry. For example, for  $\text{NH}_3$ , we would input different values for each bond length and different values for each bond angle, and would make sure no symmetry plane was present in the starting structure. (Recall that when all four atoms are in the same plane in the input geometry, *Gaussian* converges to a planar geometry.) Eliminating symmetry increases the time needed for the geometry optimization.

As noted in Section 15.12, a frequency calculation should be done at a minimum in the PES, so the frequency calculation should be preceded by a geometry optimization that uses the same method and basis set as the frequency calculation. Putting the two keywords Opt and Freq on the input line that begins with a # will cause *Gaussian* to follow the optimization with a frequency calculation.

### Output

We now consider a few points about the output of a *Gaussian* calculation. Fuller details are given in *Foresman and Frisch* and in the online *Gaussian* manual ([www.gaussian.com](http://www.gaussian.com)). The CCCBDB (Section 15.1) contains the input and output files of many *Gaussian* calculations; these can be downloaded and viewed in a word processor. You can also view *Gaussian* output files of jobs you run in the WebMO demo program by clicking Raw output.

The default procedure is for *Gaussian* not to print out the wave function. The keyword Pop = Reg will produce the coefficients of the MO expansions in terms of the basis functions for the five highest occupied and five lowest virtual MOs. The keyword Pop = Full produces all the MOs.

In a frequency calculation, *Gaussian* lists the harmonic frequencies (really the wavenumbers) in order of increasing value and in units of  $\text{cm}^{-1}$ . Any imaginary frequencies

are listed first and are preceded by a minus sign as a signal that these are imaginary frequencies. Also listed are the symmetry species (Section 15.2) and relative coordinate displacements for the normal mode corresponding to each frequency.

### Automatic Model Builders

An **automatic model builder** program is one that given the two-dimensional structure of a molecule (the atoms and their bonding linkages and specification of such stereochemical relations as cis or trans at each double bond) will attempt to construct a three-dimensional low-energy conformer without any user intervention. Comparison of the three-dimensional structures produced by several such programs with 639 accurately known organic structures found that CORINA (from the word *coordinates*) gave the best overall results [J. Gasteiger et al., *J. Chem. Inf. Comput. Sci.*, **36**, 1030 (1996); J. Sadowski, J. Gasteiger, and G. Klebe, *ibid.*, **34**, 1000 (1994); J. Sadowski and J. Gasteiger, *Chem. Rev.*, **93**, 2567 (1993)]. For 42% of the 639 structures, CORINA gave rms differences between the calculated and experimental atomic coordinates of less than 0.3 Å, and for 50% of the structures, the CORINA dihedral angles had an rms deviation from the true values of less than 15°. CORINA uses standard bond lengths and angles, assigns dihedral angles to minimize steric repulsions, and uses a simplified molecular-mechanics-like pseudo force field to do a geometry optimization for each ring in the molecule.

You can use the Internet to generate CORINA structures at no charge. Go to [www.molecular-networks.com/online\\_demos/corina\\_demo.html](http://www.molecular-networks.com/online_demos/corina_demo.html), where you can either draw the molecule you are interested in or enter its structural formula as a SMILES string. SMILES is a way of representing a structural formula by a one-dimensional string. Some examples are CC=O for CH3CHO (an equals sign denotes a double bond, and hydrogens are omitted from organic compounds), CC(C)C for (CH3)2CHCH3, and C1CCC1 for cyclobutane (the numeral ones indicate that the carbons preceding them are connected to each other). Further information on SMILES can be found at [www.daylight.com/dayhtml/doc/theory/theory.smiles.html](http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html) or in D. Weininger, *J. Chem. Inf. Comput. Sci.*, **28**, 31 (1988). After the SMILES string and the molecule's name are entered, click on Submit to produce a rotatable three-dimensional model of the CORINA structure (provided you have a molecule-viewer such as Jmol ([jmol.sourceforge.net](http://jmol.sourceforge.net)), RASMOL, or CHIME installed on your computer). If you are using the Jmol viewer, by right-clicking the structure and choosing Show and then Extract MOL data, you will get a window with the Cartesian coordinates of the atoms.

Unlike CORINA, which produces a single conformer, the program OMEGA produces up to a few hundred low-energy conformers [[www.eyesopen.com/omega](http://www.eyesopen.com/omega); P. C. D. Hawkins et al., *J. Chem. Inf. Model.*, **50**, 572 (2010)].

The free online database ChemDB [[cdb.ics.uci.edu](http://cdb.ics.uci.edu) and J. Chen et al., *Bioinformatics*, **21**, 4133 (2005); **23**, 2348 (2007)] contains about 5 million commercially available compounds whose three-dimensional structures have been estimated using CORINA; the database can be efficiently searched for desired structural features. After entering the name or SMILES string of what you are looking for, the search may give you more than one result. Choose the molecule you want by clicking on the number near its formula. You will get a rotatable three-dimensional model. In the drop-down list, choose the XYZ – Xmol XYZ format and click on Download Chemical. Save it to your computer. When you open the saved file with a word processor, you will see the Cartesian coordinates.

People doing drug design are keenly interested in finding structures whose shape fits a target receptor site and that have certain desired chemical properties that will enhance binding to the site. [For a review of protein–ligand docking, see B. Waszkowycz et al., *WIREs Comput. Mol. Sci.*, **1**, 229 (2011).] One way to find such structures is to search databases of structures. The Cambridge Structural Database [[www.ccdc.cam.ac.uk/](http://www.ccdc.cam.ac.uk/); F. H. Allen, *Acta Crystallogr. B*, **58**, 380 (2002)] contains over 600000 X-ray and neutron-diffraction-determined structures of

small- and medium-size (typically 20 to 100 atoms) organic and organometallic compounds. Databases containing millions of three-dimensional structures produced by automatic model builders exist. ChemDB was mentioned above. Another such searchable free database is ZINC [zinc.docking.org; J. J. Irwin et al., *J. Chem. Inf. Model.*, **52**, 1757 (2012)] with over 20 million commercially available compounds.

For methods of searching structural databases for desired features, see *Leach*, Chapter 12; Y. C. Martin, M. G. Bures, and P. Willett, in K. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry*, Vol. 1, VCH, Chapter 6.

## 15.16 Speeding Up Hartree–Fock Calculations

### Dealing with Electron-Repulsion Integrals

The calculation of the approximately  $b^4/8$  **electron-repulsion integrals** (ERIs)  $(rs|tu)$  [Eq. (14.39)] over the  $b$  basis functions consumes a major part of the time in an SCF MO calculation. Several methods are used to reduce the number of integrals evaluated.

Molecular symmetry is used to identify integrals that are equal, so that only one of them need be evaluated. For example, in  $\text{H}_2\text{O}$ , the integrals  $(\text{H}_1 1s \text{ O} 2s | \text{H}_2 1s \text{ H}_2 1s)$  and  $(\text{H}_2 1s \text{ O} 2s | \text{H}_1 1s \text{ H}_1 1s)$  are equal, provided the  $\text{O}-\text{H}_1$  and  $\text{O}-\text{H}_2$  bond distances are equal. Use of symmetry cuts the number of integrals to be evaluated in  $\text{H}_2\text{O}$  approximately in half.

In a large molecule, any one atom is far from most of the other atoms, and so a large fraction of the two-electron integrals are negligibly small for large molecules;  $(rs|tu)$  will be very small if  $\chi_r(1)$  and  $\chi_s(1)$  or  $\chi_t(2)$  and  $\chi_u(2)$  are centered on widely separated nuclei. Hence, many programs test each  $(rs|tu)$  integral to get its order of magnitude before it is calculated accurately. Integrals smaller than a certain threshold value can be neglected without affecting the accuracy of the overall calculation. Although the number of two-electron integrals increases as  $b^4$ , the number of such integrals whose value exceeds a fixed threshold increases only as  $b^2$  for large molecules. Calculations on large molecules done with integrals of value less than  $10^{-9}$  hartree neglected showed that the time to do an SCF calculation increased only as  $b^{2.3}$  [R. Ahlrichs et al., *Chem. Phys. Lett.*, **162**, 165 (1989); see also D. L. Strout and G. E. Scuseria, *J. Chem. Phys.*, **102**, 8448 (1995)].

Many  $(rs|tu)$  integrals involve basis functions representing inner-shell orbitals. These orbitals are little changed on molecule formation, and one can eliminate the need to explicitly represent them by using an **effective core potential** (ECP) or **pseudopotential** (Section 13.17). The ECP is a one-electron operator that replaces those two-electron Coulomb and exchange operators in the valence-electrons' Hartree–Fock equation  $\hat{F}\phi_i = \epsilon_i\phi_i$  that arise from interactions between the core electrons and the valence electrons. ECPs are derived from ab initio all-electron calculations on atoms. For compounds of main-group elements, calculations that use properly chosen ECPs give almost the same results as comparable all-electron ab initio calculations. For transition elements, obtaining accurate results with ECPs is harder. ECPs are reviewed in M. Krauss and W. J. Stevens, *Ann. Rev. Phys. Chem.*, **35**, 357 (1984); G. Frenking et al., in K. B. Lipkowitz and D. B. Boyd, eds., *Reviews in Computational Chemistry*, Vol. 8, Wiley, 1996, Chapter 2; T. R. Cundari et al., in *Reviews in Computational Chemistry*, Vol. 8, Chapter 3; M. Dolg and X. Cao, *Chem. Rev.*, **112**, 403 (2012).

The integrals  $(rs|tu)$  not only must be calculated but also must be stored and then recalled from memory as their values are needed in each SCF iteration (recall the SCF example in Section 14.3). Typically, 5 to 50 iterations are needed to achieve SCF convergence. For the large basis sets used in modern ab initio calculations, the number of  $(rs|tu)$  values to be stored for a large-molecule calculation may exceed the internal (core) memory capacity of the computer, and the  $(rs|tu)$  values must be stored on external memory disk

drives. Locating and reading in the value of an integral from external memory is a relatively slow process.

To avoid the use of external storage memory, Almlöf developed the **direct SCF method** (not to be confused with the direct CI method of Section 16.2), in which no  $(rs|tu)$  integrals are stored, but each two-electron integral is recomputed each time its value is needed. The direct SCF method allows ab initio calculations of large molecules and is very widely used.

When the direct SCF method was developed in the early 1980s, the internal memory of computers was rather limited. Nowadays, personal computers with internal memories of 16 gigabytes, hard-drive memories of a couple of thousand gigabytes, and very rapid speeds for reading and writing hard-drive data are common. This has made conventional calculations with all the integrals being stored more competitive with the direct SCF method, especially when several personal computers running in parallel are used [A. V. Mitin et al., *J. Comput. Chem.*, **24**, 154 (2003)].

A compromise between the direct method, which recomputes all integrals as needed, and the **conventional** stored-integral method, is the widely used **semidirect method**, which stores some integrals (those most time-consuming to calculate) and recomputes others.

### Rapid Evaluation of Fock Matrix Elements

The expression (14.41) for the Fock matrix elements  $F_{rs}$  in Hartree–Fock theory for a closed-shell molecule can be written as

$$F_{rs} = \langle \chi_r | \hat{F} | \chi_s \rangle = H_{rs}^{\text{core}} + J_{rs} - \frac{1}{2} K_{rs}$$

$$J_{rs} \equiv \sum_{t=1}^b \sum_{u=1}^b P_{tu}(rs|tu), \quad K_{rs} \equiv \sum_{t=1}^b \sum_{u=1}^b P_{tu}(ru|ts) \quad (15.79)$$

where the **Coulomb matrix element**  $J_{rs}$  and the **exchange matrix element**  $K_{rs}$  (which involve the basis functions  $\chi_r$  and the density matrix elements  $P_{tu}$ ) should not be confused with the Coulomb and exchange integrals  $J_{ij}$  and  $K_{ij}$  in (14.24) (which involve the MOs  $\phi_i$  and  $\phi_j$ ). Computation of the  $F_{rs}$  values is time-consuming because of the huge number of electron-repulsion integrals  $(rs|tu)$  that occur in  $J_{rs}$  and  $K_{rs}$ . To speed up calculation of the  $J_{rs}$  matrix elements, quantum chemists use ideas developed by the mathematicians Greengard and Rokhlin in their **fast multipole method** (FMM). The FMM speeds up the calculation of the potential energy  $C \sum_i \sum_{j>i} Q_i Q_j / r_{ij}$  of a system of  $N$  point charges (or point masses), when  $N$  is very large. Here,  $Q_i$  and  $Q_j$  are charges (or masses) separated by a distance  $r_{ij}$  and  $C$  is a constant. The potential energy contains  $\frac{1}{2}N(N-1)$  terms. The time required to calculate the potential energy is essentially proportional to  $N^2$  when  $N$  is large, and we say that this is an  $O(N^2)$  calculation, where  $O$  stands for “order of.”

In the FMM, the system of charges is imagined to be located in a box that is divided and subdivided repeatedly into various levels of smaller boxes. Interactions between charges close to each other (the *near field*) are calculated directly by the usual summation formula. Interactions between charges far from one another (the *far field*) are calculated using a multipole expansion. Given a group of charges close to one another, a **multipole expansion** expresses the contribution of the charge group to the potential energy at a point  $P$  outside the group as an infinite series. Each term in the series depends on the distance of the point  $P$  from the group of charges, a spherical harmonic function, and a multipole moment of the charge group. The multipole moments in successive terms of the series are the electric monopole moment (which equals the net charge of the group of charges), the electric dipole moment, the electric quadrupole moment, and so on. Accurate results can be obtained by including only a limited number of terms in the multipole expansions. If  $N$  is very large, the FMM method reduces the calculation from  $O(N^2)$  to  $O(N)$ . A related

method for dealing with this classical-physics problem is the **tree code** method of Barnes and Hut, which also uses boxes and multipole expansions but differs in calculational details from the FMM. For discussion of the FMM and tree-code methods, see L. Greengard, *Science*, **265**, 909 (1994).

Calculation of the Coulomb matrix element  $J_{rs}$  in (15.79) involves not point charges (as in the FMM method) but continuous distributions of charge defined by the basis functions. Therefore, quantum chemists modified the FMM method to deal with interactions involving continuous charge distributions. One such modification for rapid evaluation of the Coulomb matrix elements for large molecules is the **continuous fast multipole method** (CFMM) [C. A. White et al., *Chem. Phys. Lett.*, **253**, 268 (1996)]. Another is the Gaussian very fast multipole method (GvFMM) [M. C. Strain, G. E. Scuseria, and M. J. Frisch, *Science*, **271**, 51 (1996)].

The **quantum-chemical tree code** (QCTC) [M. Challacombe and E. Schwegler, *J. Chem. Phys.*, **106**, 5526 (1997)] is a modification of the classical tree-code method. The QCTC method allows calculation of the matrix elements  $J_{rs}$  of the Coulomb matrix **J** for large molecules in a time that is proportional to the number of basis functions  $b$ ; this calculation is  $O(b)$ , and one says that the calculation exhibits **linear scaling** with size of the molecule. Challacombe and Schwegler used the QCTC method to do an ab initio SCF MO calculation on the 698-atom monomer of the P53 protein at a fixed geometry (obtained from a protein data bank) using the 3-21G basis set (3836 basis functions). They then calculated the molecular electrostatic potential (Section 15.7) of the P53 monomer. (The P53 protein is a tetramer and acts as a tumor suppressor. Mutations in the gene for this protein are found in half of human cancers.)

For small molecules, evaluation of **J** by conventional methods is faster than by a fast multipole method. The number of basis functions at which the FMM becomes faster is called the *crossover point* or the *onset point*.

When a fast-multipole method is used, short-range interelectronic interactions must still be computed by evaluating the relevant electron-repulsion integrals. These integrals are tested and those whose values are negligible are omitted, but this still leaves a lot of integrals to be evaluated for a large molecule. The electron-repulsion integrals  $(rs|tu)$  are used to compute the Coulomb and exchange matrix elements  $J_{rs}$  and  $K_{rs}$ ; the values of the individual  $(rs|tu)$ 's are not in themselves really needed. To speed up evaluation of the  $J_{rs}$  values, White and Head-Gordon devised what they call a *J-matrix engine*, which is an efficient procedure that calculates the  $J_{rs}$  values without explicitly forming all the  $(rs|tu)$  integrals [C. A. White and M. Head-Gordon, *J. Chem. Phys.*, **104**, 2620 (1996)]. In the traditional approach, the  $(rs|tu)$  integrals are calculated as intermediate quantities, which are then used to calculate the  $J_{rs}$  values. The *J-matrix engine* uses a different set of intermediate quantities to calculate the  $J_{rs}$ 's, so as to reduce the computational time without making any approximations.

A widely used alternative to multipole methods to speed up the evaluation of **J** is the **resolution of the identity (RI) approximation** (also called the **density fitting method**). Here, one uses an **auxiliary basis set**  $g_k$  to expand the products  $\chi_r(1)\chi_s(1)$  and  $\chi_t(2)\chi_u(2)$  in the  $(rs|tu)$  integral:

$$\chi_r(1)\chi_s(1) = \sum_k d_{rs,k} g_k(1), \quad \chi_t(2)\chi_u(2) = \sum_k d_{tu,k} g_k(2) \quad (15.80)$$

The  $d_{rs}$  coefficients that give good fits to the  $\chi_r(1)\chi_s(1)$  products are computed by the program. The  $g_k$  set usually is atom-centered Gaussians and has 2 to 4 times the number of basis functions as the  $\chi_s$  set. Because the  $g_k$  basis set is incomplete, (15.80) is an approximation, and one must keep the errors introduced by the RI approximation small. The

expansions (15.80) allow the very numerous four-center integrals  $(rs|tu)$  to be expressed as sums containing less numerous three-center integrals involving the  $g_k$  functions and the  $d_{rs}$  coefficients. The RI method can speed up calculation of the exchange matrix elements  $K_{rs}$  as well as the  $J_{rs}$ 's. The RI method can give huge savings in time for calculations with very large basis sets  $\chi_s$  [F. Weigend, *Phys. Chem. Chem. Phys.*, **4**, 4285 (2002); C. Hättig, *ibid.*, **7**, 59 (2005)].

An alternative method to the RI approximation is the **Cholesky decomposition (CD) method**, which also uses an auxiliary basis set to evaluate electron-repulsion integrals [F. Aquilante et al., Chapter 13 in J. Leszczyński (ed.) *Linear Scaling Techniques in Computational Chemistry*, Springer, 2011; available at [uu.diva-portal.org/smash/get/diva2:396223/FULLTEXT01](http://uu.diva-portal.org/smash/get/diva2:396223/FULLTEXT01).] A comparison found the RI method to be significantly faster than the CD method but noted that Cholesky decomposition is more useful in high-accuracy benchmark calculations with very large basis sets [F. Wiegand, *J. Chem. Phys.*, **130**, 164106 (2009)].

The **multiplicative integral approximation (MIA)** is similar to the RI approximation in that it uses the expansions (15.80) to evaluate the  $(rs|tu)$ 's, but differs in details. Using the MIA method, Van Alsenoy and co-workers did an ab initio SCF MO geometry optimization (Section 15.10) of the 642-atom molecule crambin (a 46-residue protein) using the 4-21G basis set with starting coordinates taken as the structure found from X-ray crystallography [C. Van Alsenoy et al., *J. Phys. Chem. A*, **102**, 2246 (1998)]. At the time it was published, this was the largest ab initio geometry optimization ever done, although the energy minimum was not located as precisely as is done with smaller molecules. The calculation took 6300 hours (260 days) of CPU time on a workstation.

The *multipole accelerated resolution of the identity (MARI-J) approximation* finds the **J** matrix using a multipole method to evaluate the far-field ERIs and the RI method to calculate the near-field integrals [M. Sierka et al., *J. Chem. Phys.*, **118**, 9136 (2003)].

Still another method to speed up evaluation of **J** is the **Fourier-transform Coulomb (FTC) method**, which replaces the evaluation of the contributions of ERIs that involve three and four Gaussian basis functions with a numerical evaluation using a plane-wave basis set. A **plane-wave function** has the form  $e^{i(kx+ly+mz)}$ , where  $k$ ,  $l$ , and  $m$  are constants. Combining the FTC, CFMM, and *J*-engine methods in one program gives a very rapid, accurate method for computing **J** for large molecules [L. Füsti-Molnár and J. Kong, *J. Chem. Phys.*, **122**, 074108 (2005)].

The ONX (order- $N$  exchange) method and the LinK (linear exchange K) method use procedures to eliminate calculation of exchange matrix elements  $K_{rs}$  that are negligibly small and allow linear scaling to be achieved in the calculation of the exchange matrix **K** for large molecules that have a nonnegligible energy gap between the highest occupied and lowest vacant MO [E. Schwegler, M. Challacombe, and M. Head-Gordon, *J. Chem. Phys.*, **106**, 9703 (1997); C. Ochsenfeld, C. A. White, and M. Head-Gordon, *ibid.*, **109**, 1663 (1998)].

## 15.17 Solvent Effects

So far, we have treated the stationary-state quantum mechanics of an isolated molecule. The molecular properties so calculated are appropriate for gas-phase molecules not at high pressure. However, most of chemistry and biochemistry occurs in solution, and the solvent can have a major effect on the position of chemical equilibrium and on reaction rates. (For a survey of solvent effects on rates, equilibria, IR, UV, and NMR spectra, see C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry*, VCH, 1988.) We now examine solvent effects on molecular and thermodynamic properties. (See also Section 17.6 for semiempirical solvation methods.)

Consider a dilute solution of a polar solute molecule  $M$  in a polar solvent  $S$ ; for example, a solution of  $\text{CH}_3\text{Cl}$  in water. The water molecules near the  $\text{CCl}$  side of a solute molecule will tend to be oriented with their positively charged hydrogen atoms toward the negatively charged  $\text{Cl}$  atom, while water molecules on the  $\text{H}_3\text{C}$  side of a solute molecule will tend to be oriented with their negatively charged  $\text{O}$  atoms toward the methyl group. In addition, the dipole moment of a solute molecule will induce a dipole moment in each nearby solvent molecule that adds to the permanent dipole moment. The net result of these orientation and induction effects is that the solvent acquires a bulk polarization in the region of each solute molecule. The polarized solvent generates an electric field, called the **reaction field**, at each solute molecule. The reaction field distorts the solute's molecular electronic wave function from what it was in the gas phase, thereby producing an induced dipole moment that adds to the permanent gas-phase dipole moment of  $M$ . The increased dipole moment of  $M$  further polarizes the solvent, and so on.

Because of the additional dipole moment induced by the solvent's reaction field, a polar molecule will have a larger dipole moment in a polar solvent than in the gas phase. Moreover, the dipole moment of a solute molecule will fluctuate with time, as the orientations of the nearby solvent molecules fluctuate. For example, for a solute water molecule in the solvent water (that is, for pure water), a molecular dynamics simulation in which the intermolecular interactions are modeled by placing fluctuating positive point charges on each  $\text{H}$  atom, and a fluctuating negative charge on each molecular  $C_2$  axis near each  $\text{O}$  atom, and including a Lennard-Jones 6–12 potential interaction between each pair of  $\text{O}$  atoms (the TIP4P-FQ model, standing for transferable intermolecular potential with four interaction sites and fluctuating charges) gives the average electric dipole moment  $\langle \mu \rangle$  at  $25^\circ\text{C}$  as 2.6 D (compared with 1.85 D in the gas phase) and gives a distribution of dipole moments with a full width at half maximum of 0.5 D [S. W. Rick et al., *J. Chem. Phys.*, **101**, 6141 (1994)].  $\langle \mu \rangle$  of liquid water cannot be directly measured (and its meaning is subject to some ambiguity, since the electron density between water molecules is not zero, giving rise to controversy about the best definition). However, since this simulation gives good results for the dielectric constant and since other simulations with different models give similar results for  $\langle \mu \rangle$ , we can be reasonably confident in the results for  $\langle \mu \rangle$ . Several other theoretical studies gave values in the range 2.4 to 3.1 D; see A. V. Gubskaya and P. G. Kusalik, *J. Chem. Phys.*, **117**, 5290 (2002); D. D. Kemp and M. S. Gordon, *J. Phys. Chem. A*, **112**, 4885 (2008); H. C. Georg and S. Canuto, *J. Phys. Chem. B*, **116**, 11247 (2012).

The molecular electronic wave function and all molecular properties in solution will differ to some extent from their gas-phase counterparts.

The rigorous way to deal with solvent effects on molecular properties is to carry out quantum-mechanical calculations on a system consisting of a solute molecule surrounded by many solvent molecules. One repeats the calculations for various orientations of the solvent molecules and takes a suitable average over orientations to find average properties at a particular temperature and pressure. Such a calculation is usually impractical. Calculations that include a number of individual solvent molecules are called **explicit solvent** calculations, and discussion of these is omitted (see *Cramer*, Chapter 12 for details).

Perhaps the most common way to calculate solvent effects is to use a **continuum solvent model**. Here, the molecular structure of the solvent is ignored, and the solvent is modeled as a continuous dielectric of infinite extent that surrounds a cavity containing the solute molecule  $M$ . (A dielectric is a nonconductor of electricity.) The continuous dielectric is characterized by its *dielectric constant* (also called *relative permittivity*)  $\epsilon_r$ , whose value is the experimental dielectric constant of the solvent at the temperature and pressure of the solution. The solute molecule can be treated classically as a collection of charges that interacts with the dielectric, or it can be treated quantum mechanically. In a quantum-mechanical treatment, the interaction between a solute molecule  $M$  and the surrounding

dielectric continuum is modeled by a term  $\hat{V}_{\text{int}}$  that is added to the molecular electronic (fixed-nuclei) Hamiltonian  $\hat{H}_{\text{M}}^{(0)}$ , where  $\hat{H}_{\text{M}}^{(0)}$  is for M in vacuum.

For reviews of continuum solvation models, see *Cramer*, Chap. 11; C. J. Cramer and D. G. Truhlar, *Chem. Rev.*, **99**, 2161 (1999); F. J. Luque et al., *Phys. Chem. Chem. Phys.*, **5**, 3827 (2003).

In the usual quantum-mechanical implementation of the continuum solvation model, the electronic wave function and electronic probability density of the solute molecule M are allowed to change upon going from the gas phase to the solution phase, so as to achieve self-consistency between the M charge distribution and the solvent's reaction field. Any treatment in which such self-consistency is achieved is called a **self-consistent reaction-field** (SCRF) model. Many versions of SCRF models exist. These differ in how they choose the size and shape of the cavity that contains the solute molecule M and in how they calculate  $\hat{V}_{\text{int}}$ .

### The Quantum-Onsager SCRF Method

In the **dipole-in-a-sphere** (or **quantum-Onsager** or Born–Kirkwood–Onsager) SCRF method, the molecular cavity is a sphere of radius  $a$  and the interaction between the molecular charge distribution and the reaction field is calculated by approximating the molecular charge distribution as an electric dipole located at the cavity center with electric dipole moment  $\boldsymbol{\mu}$ . In 1936, Onsager showed that the electric field in the cavity (the reaction field) produced by the polarization of the solvent by  $\boldsymbol{\mu}$  is (in atomic units)

$$\mathbf{E}_R = \frac{2(\epsilon_r - 1)}{(2\epsilon_r + 1)a^3} \boldsymbol{\mu} \quad (15.81)$$

The potential energy of electrostatic interaction between  $\boldsymbol{\mu}$  and the reaction field  $\mathbf{E}_R$  is  $\hat{V}_{\text{int}} = -\boldsymbol{\mu} \cdot \mathbf{E}_R$ . The corresponding quantum-mechanical operator in atomic units is

$$\hat{V}_{\text{int}} = -\hat{\boldsymbol{\mu}} \cdot \mathbf{E}_R, \quad \hat{\boldsymbol{\mu}} = -\sum_i \mathbf{r}_i + \sum_{\alpha} Z_{\alpha} \mathbf{R}_{\alpha} \quad (15.82)$$

where the electric dipole moment operator was taken from (14.20).

In an SCRF quantum-Onsager calculation, one starts by using a method such as HF, DFT (Section 16.5), MP2 (Section 16.3), or whatever, to calculate an electron probability density  $\rho^{(0)}(\mathbf{r})$  for the isolated molecule, preferably at an optimized geometry. [ $\rho^{(0)}$  is calculated from (14.5) in a wave-function-based method or from (16.45) in density-functional theory.] One then calculates the electric dipole moment in vacuum from (14.21) as  $\boldsymbol{\mu}^{(0)} = -\int \rho^{(0)}(\mathbf{r}) \mathbf{r} d\mathbf{r} + \sum_{\alpha} Z_{\alpha} \mathbf{R}_{\alpha}$ . Then  $\boldsymbol{\mu}^{(0)}$  is used in (15.81) to give an initial estimate  $\mathbf{E}_R^{(0)}$  of the reaction field. From  $\mathbf{E}_R^{(0)}$ , one calculates an initial estimate of the operator  $\hat{V}_{\text{int}}$  as  $\hat{V}_{\text{int}}^{(0)} = -\hat{\boldsymbol{\mu}} \cdot \mathbf{E}_R^{(0)}$ , where  $\hat{\boldsymbol{\mu}}$  is given by (15.82). Using  $\hat{V}_{\text{int}}^{(0)}$ , one solves the equations of the quantum method being used and obtains an improved electron probability density  $\rho^{(1)}$ .

For example, suppose the Hartree–Fock method is being used. One finds that the part of  $\hat{V}_{\text{int}}$  in (15.82) that involves electron 1 is added to the Fock operator  $\hat{F}(1)$  of the Hartree–Fock equations (14.25) to give as the Hartree–Fock equations

$$[\hat{F}(1) + \mathbf{r}_1 \cdot \mathbf{E}_R] \phi_i(1) = \epsilon_i \phi_i(1)$$

for the orbitals in the presence of the reaction field. Since the nuclei are fixed in solving the electronic Schrödinger equation, the interaction  $-\sum_{\alpha} Z_{\alpha} \mathbf{R}_{\alpha} \cdot \mathbf{E}_R$  of the nuclear charges with the reaction field is a constant term, that, like  $V_{NN}$  in (13.6), appears in the electronic energy but not in the equations for the orbitals. [In density-functional theory (Section 16.5), the Kohn–Sham operator is similarly modified.]

From  $\rho^{(1)}$ , one calculates from (14.21) a value  $\boldsymbol{\mu}^{(1)}$  for the molecular dipole moment that allows for the effect of  $\mathbf{E}_R^{(0)}$ . Using  $\boldsymbol{\mu}^{(1)}$  in (15.79), one obtains an improved estimate

$\mathbf{E}_R^{(1)}$  for the reaction field. Then an improved  $\hat{V}_{\text{int}}^{(1)} = -\hat{\boldsymbol{\mu}} \cdot \mathbf{E}_R^{(1)}$  is used to solve for an improved electron density. And so on. Iterations are continued until there is no further change in  $\rho$ ,  $\boldsymbol{\mu}$ , and  $\mathbf{E}_R$ . The molecular geometry should be reoptimized in the presence of the reaction field, but this step is sometimes omitted, since changes in geometry from the gas phase to the solution phase are usually small.

If the solute species has a net charge  $Q$ , the constant term  $-(\epsilon_r - 1)Q^2/\epsilon_r a$  (first derived by Born), is added to  $V_{\text{int}}$ .

In the quantum-Onsager SCRF method, an uncharged solute molecule with no permanent dipole moment is unaffected by the solvent. In reality, the quadrupole and higher moments of the solute will interact with the solvent to give a reaction field.

With  $\hat{V}_{\text{int}}$  included in the Hamiltonian operator, one obtains a molecular electronic energy (including nuclear repulsion)  $U^{(f)}$  that is an eigenvalue of  $\hat{H}_M^{(0)} + \hat{V}_{\text{int}}$ . We have  $U^{(f)} = \langle \psi^{(f)} | \hat{H}_M^{(0)} + \hat{V}_{\text{int}} | \psi^{(f)} \rangle$ , where  $\psi^{(f)}$  is the final electronic wave function of M, found when self-consistency is attained. The molecular electronic energy in vacuum is  $U^{(0)} = \langle \psi^{(0)} | \hat{H}_M^{(0)} | \psi^{(0)} \rangle$ . Thus

$$U^{(f)} = \langle \psi^{(f)} | \hat{H}_M^{(0)} + \hat{V}_{\text{int}} | \psi^{(f)} \rangle, \quad U^{(0)} = \langle \psi^{(0)} | \hat{H}_M^{(0)} | \psi^{(0)} \rangle \quad (15.83)$$

The energy  $U^{(f)}$  includes the energy of solute interaction with the solvent-produced reaction field, but there is another energy contribution that must also be included. This is the energy change in the solvent that results from its being polarized by the solute. The work required to polarize the solvent can be shown to be

$$E_{\text{pol}} = -\frac{1}{2} \langle \psi^{(f)} | \hat{V}_{\text{int}} | \psi^{(f)} \rangle \quad (15.84)$$

### Gibbs Energy of Solvation

The standard Gibbs free-energy change  $\Delta G_{\text{solv}}^\circ$  for solvation of the solute species M in the solvent S at a particular temperature is defined as  $\Delta G^\circ$  for the process of gas-phase M plus liquid-phase S going to a solution of M in S. The standard states most commonly used in the definition of  $\Delta G_{\text{solv}}^\circ$  are as follows. The solute standard state is the fictitious state with the solute at a 1 mole per liter concentration in the solvent but with the solute experiencing only interactions with solvent molecules and not with other solute molecules (as would be true in an ideally dilute solution). The gas-phase standard state of M is the fictitious state of pure gaseous solute at a 1 mole per liter concentration but with no intermolecular interactions (as would be true in an ideal gas). These standard states are not the ones most commonly used in thermodynamics. Confusion exists in the literature since people dealing with solvation energies often do not specify what standard states they are using.

Theoretical treatments of  $\Delta G_{\text{solv}}^\circ$  separate it into several components. The **electrostatic contribution**  $\Delta G_{\text{solv,el}}^\circ$  to  $\Delta G_{\text{solv}}^\circ$  results from the electrostatic interactions between the solute and solvent. One can show that  $\Delta G_{\text{solv,el}}^\circ$  can be found from an SCRF calculation as follows:

$$\begin{aligned} \Delta G_{\text{solv,el}}^\circ &= (U^{(f)} + E_{\text{pol}}) - U^{(0)} = U^{(f)} - \frac{1}{2} \langle \psi^{(f)} | \hat{V}_{\text{int}} | \psi^{(f)} \rangle - U^{(0)} \\ \Delta G_{\text{solv,el}}^\circ &= \langle \psi^{(f)} | \hat{H}_M^{(0)} + \frac{1}{2} \hat{V}_{\text{int}} | \psi^{(f)} \rangle - \langle \psi^{(0)} | \hat{H}_M^{(0)} | \psi^{(0)} \rangle \end{aligned} \quad (15.85)$$

(Since  $\Delta G_{\text{solv}}^\circ$  is expressed on a per-mole basis, the right sides of these equations should be multiplied by the Avogadro constant  $N_A$ .) Although energy is both a molecular and a macroscopic property, both entropy and free energy are macroscopic but not molecular properties. The dielectric continuum model uses the macroscopic property  $\epsilon_r$  of the solvent, and the macroscopic treatment of the solvent allows us to find  $\Delta G_{\text{solv,el}}^\circ$ , which is a contribution to a macroscopic property. The dielectric continuum treatment of solvation is a combined quantum-mechanical and statistical-mechanical treatment.

In addition to the electrostatic contribution to  $\Delta G_{\text{solv}}^\circ$ , there are the following contributions. The **cavitation** contribution  $\Delta G_{\text{solv,cav}}^\circ$  is the work needed to form the cavities in the solvent that are occupied by the solute molecules. The **dispersion** contribution  $\Delta G_{\text{solv,dis}}^\circ$  results from London dispersion attractions between solute and solvent molecules. The **repulsion** contribution  $\Delta G_{\text{solv,rep}}^\circ$  (often called the exchange-repulsion contribution) results from quantum-mechanical repulsions between solute and solvent molecules. The **molecular motion** (or **thermal**) contribution is  $\Delta G_{\text{solv,mm}}^\circ = -RT \ln(z_{\text{M(sln)}}/z_{\text{M(g)}})$ , where  $z_{\text{M(sln)}}$  and  $z_{\text{M(g)}}$  are the molecular partition functions of M in the solution and in the gas at the temperature  $T$ , and where 1 mol/L standard states are used in both the gas and in solution. This contribution results from changes in molecular motions upon going from the gas to the solution phase. (Recall the statistical-mechanical formula  $A - U_0 = -kT \ln Z$  for the Helmholtz free energy, where  $Z$  is the canonical partition function and  $U_0$  is the internal energy at absolute zero, with all molecules in the lowest energy level.) There is also a  $P \Delta V$  contribution resulting from the difference between  $\Delta A$  and  $\Delta G$ , which is negligible. Some workers also include a contribution from specific structural interactions, such as hydrogen bonding, between solute and solvent molecules. Much of the effect of hydrogen bonding is included in the electrostatic contribution.

Several methods have been proposed to calculate  $\Delta G_{\text{solv,cav}}^\circ$ . The *scaled-particle-theory* formula (derived by Reiss, Frisch, Helfand, and Lebowitz and first applied to a solute in solution by Pierotti) calculates  $\Delta G_{\text{solv,cav}}^\circ$  from the radii of the solute and solvent molecules (assumed spherical), the number of solvent molecules per unit volume, and the temperature and pressure [H. Reiss et al., *J. Chem. Phys.*, **32**, 119 (1960)]. For solvation methods that use a nonspherical molecular cavity, the scaled-particle theory spherical-cavity formula for  $\Delta G_{\text{solv,cav}}^\circ$  was modified by Claverie to give what is often called the Pierotti–Claverie formula. The available methods give quite different results for  $\Delta G_{\text{solv,cav}}^\circ$ ; which method is best is unclear, since  $\Delta G_{\text{solv,cav}}^\circ$  is not a measurable quantity. Monte Carlo simulations of liquid water indicate that the Pierotti–Claverie formula may be significantly in error [F. M. Floris et al., *J. Chem. Phys.*, **107**, 6353 (1997)]. For details, see J. Tomasi and M. Persico, *Chem. Rev.*, **94**, 2027 (1994), Section V.B.

The dispersion energy is commonly approximated as the sum of dispersion attractions between each atom of the solute molecule M and each atom of the surrounding solvent molecules, where each atom–atom dispersion energy has the form [Eq. (14.103)]  $-d_{ij}/r_{ij}^6$ , where  $r_{ij}$  is the distance between atoms  $i$  and  $j$ , and the coefficient  $d_{ij}$  is taken from experimental data or from theoretical calculations. The dispersion energy is proportional to the density of solvent molecules and involves the distribution function that gives the probability density for atoms  $i$  and  $j$  being a distance  $r_{ij}$  apart. A commonly used approximation for this probability density is to take it as zero when the solvent atom  $j$  is in the molecular cavity of M and to take it as a constant otherwise. Once the solute–solvent dispersion energy is obtained, it must be converted to a dispersion free energy by a calculational procedure (see Tomasi and Persico, op. cit., Section V.C).

The repulsion contribution  $\Delta G_{\text{solv,rep}}^\circ$  is calculated the same way as the dispersion contribution, except that  $-d_{ij}/r^6$  is replaced by  $c_{ij}/r^{12}$  [recall the Lennard-Jones 6–12 potential—Eq. (17.89)]. The repulsion contribution is much smaller than the dispersion contribution and is often omitted.

In the molecular-motion contribution, the molecular partition function is the product of translational, rotational, vibrational, and electronic partition functions. If the molecule in solution is assumed to have the entire volume of the solution available to it, the ratio of gas-phase and solution-phase translational partition functions equals one. Likewise, the electronic partition function ratio will be one. It is unclear what one should use for the rotational partition function in solution, but if this is assumed to have the same form as that in the gas phase, the rotational partition function ratio (which involves the moments of

inertia) will be very close to one, since structural changes from gas to solution are slight. Significant contributions to the vibrational partition function are made only by the low-frequency vibrational normal modes, and these modes sometimes show substantial changes in frequency on going from the gas phase to solution. If a vibrational calculation is done in the gas phase and in solution, one can calculate  $\Delta G_{\text{solv,mm}}^\circ$ , but the most common procedure is to omit it, assuming that its contribution is negligible.

### The Multipole-Expansion Method

Two crude approximations in the quantum-Onsager SCRF method are the use of a spherical cavity for the solute molecular shape and the replacement of the actual distribution of solute molecular charge by a dipole. The true potential energy of interaction between a distribution of molecular charge and the surrounding continuum dielectric can be written as an infinite series (called a **multipole expansion**), in which the first term for a neutral molecule involves the molecular dipole moment [see Eq. (14.21)], the second term involves the molecular quadrupole moment, the third term involves the molecular octupole moment, and so on. Whereas the quantum-Onsager method includes only the first term, the multipole expansion method includes higher terms as well. The number of terms included is decided by the person doing the calculation. One finds that the terms after the Onsager dipole term make substantial contributions, and neglect of these terms is not justified. The fact that dipole-in-a-sphere SCRF calculations sometimes give good results has been ascribed to a partial cancellation of errors, with the neglect of the higher electrostatic terms partly canceling the error caused by using a spherical molecular shape [J. B. Foresman et al., *J. Phys. Chem.*, **100**, 16098 (1996)].

An improvement on a spherical molecular shape is an ellipsoidal molecular shape. Quantum-Onsager and multipole-expansion SCRF calculations using an ellipsoidal cavity give better results than spherical-cavity results, but the improvement is not great.

### The PCM Method

Accurate ab initio calculation of solvent effects requires a molecular shape more realistic than a sphere or an ellipsoid. In the **polarizable-continuum model** (PCM) of Miertus, Scrocco, and Tomasi, each atomic nucleus in the solute molecule *M* is surrounded by a sphere of radius 1.2 times the van der Waals radius of that atom [see the references in V. Barone et al., *J. Comput. Chem.*, **19**, 404 (1998)]. (The volume occupied by these overlapping atomic spheres has sharp crevices where spheres intersect, so a smoothing procedure is sometimes used to eliminate the crevices.)

Since the PCM cavity has a complex shape, analytic expressions for the expansion coefficients in a multipole expansion cannot be found, and an analytic multipole expansion method is not feasible. Instead, a numerical method is used to find the solute-solvent interaction potential energy term  $\hat{V}_{\text{int}}$ . One can show from classical electrostatics that the electric potential  $\phi_\sigma$  produced by the polarized dielectric continuum is equal to the electric potential produced by an **apparent surface charge** (ASC) distributed on the surface of the molecular cavity. The ASC is a continuous distribution of charge characterized by a surface charge density (charge per unit surface area) that varies from point to point of the cavity surface. In practice, one approximates this continuous ASC by replacing it with many point charges on the cavity surface. The cavity surface is divided into many tiny regions, and an apparent charge  $Q_k$  is placed in the *k*th region. If  $\mathbf{r}_k$  is the point at which  $Q_k$  is located, then the electric potential  $\phi_\sigma(\mathbf{r})$  due to the polarization of the dielectric is (in atomic units) [Eq. (15.27)]

$$\phi_\sigma(\mathbf{r}) = \sum_k \frac{Q_k}{|\mathbf{r} - \mathbf{r}_k|} \quad (15.86)$$

Classical electrostatics gives the following expression for the apparent charges:

$$Q_k = [(\epsilon_r - 1)/4\pi\epsilon_r] A_k \nabla \phi_{\text{in}}(\mathbf{r}_k) \cdot \mathbf{n}_k \quad (15.87)$$

where  $A_k$  is the area of the  $k$ th region,  $\mathbf{r}_k$  is the point at which  $Q_k$  is located,  $\nabla \phi_{\text{in}}(\mathbf{r}_k)$  is the gradient of the electric potential within the cavity evaluated in the limit as point  $\mathbf{r}_k$  is approached, and  $\mathbf{n}_k$  is a unit vector perpendicular to the cavity surface at  $\mathbf{r}_k$  and pointing out of the cavity. [The gradient of  $\phi$  is discontinuous at the cavity surface, so it is necessary to distinguish  $\nabla \phi_{\text{in}}(\mathbf{r}_k)$  from  $\nabla \phi_{\text{out}}(\mathbf{r}_k)$ .] The electric potential within the cavity is the sum of the contribution  $\phi_{\text{M, in}}$  from the charge distribution of the solute molecule M and the contribution  $\phi_{\sigma, \text{in}}$  from the polarized dielectric:  $\phi_{\text{in}} = \phi_{\text{M, in}} + \phi_{\sigma, \text{in}}$ .

Since neither  $\phi_{\text{in}}$  nor  $Q_k$  is known initially, one finds the apparent surface charges by the following iterative process. One initially neglects  $\phi_{\sigma, \text{in}}$  and takes the initial estimate of  $\phi_{\text{in}}$  as  $\phi_{\text{in}}^{(00)} = \phi_{\text{M, in}}^{(0)}$  where  $\phi_{\text{M, in}}^{(0)}$  is calculated from the electron density  $\rho^{(0)}$  of an M molecule in vacuum. [ $\rho^{(0)}$  is found from the wave function or, in DFT, the Kohn–Sham orbitals, of M in vacuum. Note that  $\phi_{\text{M}}^{(0)}$  is the molecular electrostatic potential of Eq. (15.28).] Then (15.87) is used to find initial estimates  $Q_k^{(00)}$  of the ASCs, which are used in (15.86) to find an initial estimate  $\phi_{\sigma, \text{in}}^{(00)}$  of the electric potential produced by the polarized dielectric. The improved potential  $\phi_{\text{in}}^{(01)} = \phi_{\text{M, in}}^{(0)} + \phi_{\sigma, \text{in}}^{(00)}$  is used in (15.87) to get improved charges  $Q_k^{(01)}$ , which are used in (15.86) to find  $\phi_{\sigma, \text{in}}^{(01)}$ , and so on. (The first superscript zero in  $Q_k^{(00)}$  and  $\phi_{\sigma, \text{in}}^{(01)}$  indicates the use of  $\phi_{\text{M, in}}^{(0)}$ .) One continues iterating until the charges converge to values  $Q_k^{(0f)}$ .

The converged charges are used to find an initial estimate of  $\hat{V}_{\text{int}}$  as

$$\hat{V}_{\text{int}}^{(0)} = -\sum_i \phi_{\sigma}^{(0f)}(\mathbf{r}_i) + \sum_{\alpha} Z_{\alpha} \phi_{\sigma}^{(0f)}(\mathbf{r}_{\alpha})$$

where the sums go over the electrons and nuclei and  $\phi_{\sigma}^{(0f)}$  is found from (15.86) using the  $Q_k^{(0f)}$ 's.  $\hat{V}_{\text{int}}^{(0)}$  is added to the molecular electronic Hamiltonian, which is used to get an improved electron density  $\rho^{(1)}$  for M, which gives  $\phi_{\text{M, in}}^{(1)}$ , which gives the improved potential  $\phi_{\text{in}}^{(10)} = \phi_{\text{M, in}}^{(1)} + \phi_{\sigma, \text{in}}^{(0f)}$ , which is used in (15.87) to start a new cycle of charge and  $\phi_{\sigma}$  iterations. One continues until everything has converged.

The original PCM method uses atomic spheres with radii 1.2 times the van der Waals radii to define the molecular cavity. The **isodensity polarizable continuum model** (IPCM) is a modification of the PCM that defines the surface of the molecular cavity as a contour surface of constant electron probability density of the solute molecule M [J. B. Foresman et al., *J. Phys. Chem.*, **100**, 16098 (1996)]. The isodensity value 0.0004 electrons/bohr<sup>3</sup> is commonly used, but other values have also been recommended [C.-G. Zhan and D. M. Chipman, *J. Chem. Phys.*, **109**, 10543 (1998)]. Since the solute's electronic wave function changes in each SCRF iteration, the size of the molecular cavity changes in each IPCM iteration. In the IPCM method,  $\hat{V}_{\text{int}}$  is calculated from apparent surface charges. The **self-consistent isodensity PCM** (SCIPCM) method is a refinement of the IPCM method (Foresman and Frisch, Chapter 10), which allows geometry optimization and vibrational-frequency calculations to be done for the solute molecule in solution.

The united-atom Hartree–Fock (UAHF) PCM method uses atomic spheres to define the molecular cavity, but the assignment of sphere radii is more involved than in the original PCM method. In the UAHF PCM method [V. Barone, M. Cossi, and J. Tomasi, *J. Chem. Phys.*, **107**, 3210 (1997)], hydrogen atoms are not assigned spheres, but are included within the sphere of the atom they are bonded to (hence the name united atom). The sphere radius  $R_X$  of a nonhydrogen atom X is given by a formula that contains parameters. The method has 11 parameters, whose values were chosen so that the method gives good results for aqueous solvation free energies. (When the method is applied to a different solvent than water, different values must be used for the 11 parameters.) Using

HF/6-31G\* calculations, geometries optimized in vacuum but not reoptimized in solution, and including electrostatic, cavitation, dispersion, and repulsion contributions, the UAHF PCM method gave free energies of solvation of 43 uncharged organic molecules in water with a mean absolute error of 0.2 kcal/mol, and a maximum error of 0.6 kcal/mol, a very good result. (To give an idea of the order of magnitude of the contributions, for  $\text{CH}_3\text{NH}_2$  in water, the electrostatic contribution was  $-4.4$  kcal/mol, the cavitation contribution was 6.8 kcal/mol, and the dispersion plus repulsion contribution was  $-7.0$  kcal/mol.)

The PCM method has been reformulated to eliminate the iterative calculation of the solute's wave function in solution. In this reformulation, the mutually consistent solute wave function in solution and the interaction operator  $\hat{V}_{\text{int}}$  are found directly in a single SCF cycle, thereby speeding up the calculations [M. Cossi et al., *Chem. Phys. Lett.*, **255**, 327 (1996)].

The PCM method is also referred to as D-PCM (dielectric PCM).

The integral equation formulation PCM (IEF-PCM) method is a generalization of the PCM method that allows one to deal with anisotropic solvents such as liquid crystals, as well as with isotropic solvents [E. Cancès, B. Mennucci, and J. Tomasi, *J. Chem. Phys.*, **107**, 3032 (1997)]. IEF-PCM has given good results when applied to aqueous ionic solutions [M. Cossi et al., *Chem. Phys. Lett.*, **286**, 253 (1998)].

The COSMO (conductorlike solvation model) method of treating solvation resembles the PCM method in using a realistic solute-molecule shape and in using surface charges on the cavity surface around the solute molecule, but these charges are initially calculated using a condition suitable for a solvent medium that is an electrical conductor rather than a dielectric. The initial charges are then multiplied by the function  $(\epsilon_r - 1)/(\epsilon_r + 0.5)$  to yield approximations for the charges suitable for the dielectric solvent (A. Klamt and G. Schüürmann, *J. Chem. Soc. Perkin Trans. 2*, **1993**, 799). The simplified procedure for finding the charges make COSMO computationally fast. A particular implementation of COSMO that allows efficient geometry optimization in solution is called C-PCM (conductor PCM) [V. Barone and M. Cossi, *J. Phys. Chem. A*, **102**, 1995 (1998)].

COSMO-RS (COSMO for real solvents) is an extension of COSMO beyond the dielectric-continuum approximation [A. Klamt et al., *J. Phys. Chem. A*, **102**, 5074 (1998); A. Klamt, *COSMO-RS*, Elsevier, 2005]. COSMO-RS begins with a quantum-mechanical COSMO calculation, and follows this with a statistical-mechanical procedure whose aim is to reduce the errors introduced by the continuum-solvent approximation. The ability of COSMO-RS to predict activity coefficients, vapor-liquid equilibria, and enthalpies of mixing was tested for many solutions [H. Grensemann and J. Gmehling, *Ind. Eng. Chem. Res.*, **44**, 1610 (2005)]. COSMO-RS predictions were found to be much less reliable than those of empirical group-contribution methods traditionally used by chemical engineers. Nevertheless, the authors recommended the use of COSMO-RS for situations where the parameters needed to use group-contribution methods are not available. [See also A. Klamt, *Ind. Eng. Chem. Res.*, **51**, 13538 (2012); J. Gmehling et al., *ibid.*, **51**, 13541 (2012).] For a review of COSMO and COSMO-RS, see A. Klamt, *WIREs, Comput. Mol. Sci.*, **1**, 699 (2011).

### Chemical Equilibria in Solution

From gas-phase values of  $\Delta U_r^\circ$  and  $\Delta S_r^\circ$  for a reaction, calculated as described in Section 15.13, we can find a theoretical value of the gas-phase  $\Delta G_r^\circ$  for a reaction at the desired temperature. Then one calculates  $\Delta G_{\text{solv}}^\circ$  for each reactant and product and combines these values with the gas-phase  $\Delta G_r^\circ$  value to find  $\Delta G_r^\circ$  for the reaction in solution, which then allows calculation of the equilibrium constant in solution.

This procedure has been applied to several isomerization reactions and also to calculate relative populations of different conformers in solution. For isomerizations or conformational changes, it's usually a good approximation to assume the two species

have nearly the same cavitation and dispersion contributions, so these contributions will essentially cancel and need not be calculated.

For example, for the keto–enol equilibrium 2-pyridone  $\rightleftharpoons$  2-hydroxypyridine, dipole-in-a-sphere ab initio calculations with cavitation and dispersion contributions omitted gave the following values at 298 K in the gas phase and in the solvents cyclohexane ( $\epsilon_r = 2.0$ ) and acetonitrile ( $\epsilon_r = 36$ ):  $\Delta G_r^\circ = -0.6, 0.4$ , and  $2.3$  kcal/mol, respectively, compared with the estimated experimental values  $-0.8, 0.3$ , and  $3.0$  kcal/mol [M. W. Wong, K. B. Wiberg, and M. J. Frisch, *J. Am. Chem. Soc.*, **114**, 1645 (1992)]. The more-polar keto form 2-pyridone (calculated gas-phase  $\mu = 4.2$  D) was stabilized more than the enol form (calculated gas-phase  $\mu = 1.5$  D) in solution.

### Molecular Properties in Solution

The following data give examples of changes in calculated molecular properties from the gas phase to solution. For the gauche conformer of 1,2-dichloroethane, some SCIPCM density-functional calculated structural changes on going from the gas phase to a dilute solution in a solvent with  $\epsilon_r = 47$  are:  $r(\text{CC})$  from 1.512 to 1.505 Å,  $r(\text{CCl})$  from 1.810 to 1.820 Å,  $\angle \text{CCCl}$  from 112.9° to 112.8°,  $D(\text{CICCCl})$  from 69.9° to 68.6°, and the dipole moment changed from 2.92 to 3.82 D [K. B. Wiberg et al., *J. Phys. Chem.*, **99**, 9072 (1995)]. For formamide,  $\text{HCONH}_2$ , some PCM density-functional calculated vibrational wavenumber changes from the gas phase to the solvent water with  $\epsilon_r = 78$  are 162 to 366  $\text{cm}^{-1}$  for  $\text{NH}_2$  inversion, 650 to 682  $\text{cm}^{-1}$  for  $\text{NH}_2$  twisting, 1276 to 1302  $\text{cm}^{-1}$  for CN stretching, and 1621 to 1588  $\text{cm}^{-1}$  for  $\text{NH}_2$  bending [V. Barone, M. Cossi, and J. Tomasi, *J. Comput. Chem.*, **19**, 404 (1998)]. For  $\text{H}_2\text{O}$ , calculations and experiments give gas-to-liquid changes in bond length and angle from 0.958 Å to 0.97 Å and from 104.5° to 106° [Table 5 in H. C. Georg and S. Canuto, *J. Phys. Chem. B*, **116**, 11247 (2012)].

### Problems

| Sec.   | 15.2      | 15.4       | 15.5        | 15.6        | 15.7        | 15.9        |
|--------|-----------|------------|-------------|-------------|-------------|-------------|
| Probs. | 15.1–15.3 | 15.4–15.12 | 15.13–15.22 | 15.23–15.25 | 15.26–15.27 | 15.28–15.30 |

| Sec.   | 15.10       | 15.11       | 15.12 | 15.13       | 15.15       | 15.16 | general |
|--------|-------------|-------------|-------|-------------|-------------|-------|---------|
| Probs. | 15.31–15.35 | 15.36–15.37 | 15.38 | 15.39–15.40 | 15.41–15.60 | 15.61 | 15.62   |

- 15.1** Verify that (15.3) are the possible symmetry species for  $\mathcal{C}_{2v}$ .
- 15.2** Work out the possible symmetry species for  $\mathcal{D}_2$ .
- 15.3** How many independent molecular wave functions correspond to (a) a  ${}^3E$  term? (b) a  ${}^1E$  term?
- 15.4** (a) Give the number of CGTFs used in a  $[4s2p/2s]$  calculation of  $\text{C}_3\text{H}_7\text{OH}$ . (b) Give the number of CGTFs used in a DZ calculation of  $\text{C}_4\text{H}_9\text{OH}$ .
- 15.5** For  $\text{C}_4\text{H}_9\text{OH}$ , give the number of CGTFs used in a calculation with each of the following basis sets: (a) STO-3G; (b) 3-21G; (c) 6-31G\*; (d) 6-31G\*\*; (e) 6-31+G\*; (f) cc-pVTZ; (g) cc-pVQZ; (h) aug-cc-pVDZ.
- 15.6** How many primitive and how many contracted GTFs are used in a calculation on  $\text{Si}_{24}\text{O}_{60}\text{H}_{24}$  with each of these basis sets: (a) STO-3G; (b) 3-21G; (c) 6-31G\*?
- 15.7** Use the CCCBDB (Section 15.1) to find the electronic energy, dipole moment, and equilibrium geometry for the ground electronic state of  $\text{NH}_3$  given by a HF/6-31G\* calculation and by a HF/cc-pVDZ calculation.
- 15.8** Repeat Prob. 15.7 for the gauche and the anti conformers of butane (see Fig. 15.13). Be sure and give the CCCC dihedral angle for each. Comment on the dipole moment values.
- 15.9** Use a spreadsheet to do a least-squares fit of the STO  $S \equiv \pi^{-1/2}e^{-r}$  with a normalized linear combination  $G_{3N}$  of three normalized  $s$ -type Cartesian Gaussians. Suppose we divide space

into infinitesimal volume elements  $d\tau$ , compute  $(S - G_{3N})^2 d\tau$  at a point in each volume element, sum these quantities to give the definite integral  $I \equiv \int (S - G_{3N})^2 d\tau$ , and vary the six parameters in  $G_{3N}$  to minimize  $I$ . (a) Show that  $I = 4\pi \int_0^\infty (S - G_{3N})^2 r^2 dr$ . (b) Set up a spreadsheet to compute  $I$  as  $I \approx \sum_{i=1}^{300} [S(r_i) - G_{3N}(r_i)]^2 4\pi r_i^2 \Delta r$ , where  $r_i$  goes from 0 to 15 in steps of  $\Delta r = 0.05$ . (Beyond  $r = 15$ , the integrand in  $I$  is negligible.) Assign cells to the orbital exponents  $\alpha_1, \alpha_2, \alpha_3$ , the multiplicative coefficients  $d_1, d_2, d_3$  of the individual normalized Gaussians in the unnormalized Gaussian (which we shall call  $G_3$ ), and the coefficients  $c_1 = Nd_1, c_2 = Nd_2, c_3 = Nd_3$  in the normalized Gaussian  $G_{3N}$ . Start with initial guesses for the six parameters (the  $\alpha$ 's and the  $d$ 's). Put the  $r_i$  values in column A, the  $S(r_i)$  values in column B, the unnormalized function  $G_3$  in column C, the quantities  $[G_3(r_i)]^2 r_i^2$  in column D, the normalized function  $G_{3N} = NG_3$  in column E, and the quantities  $4\pi \Delta r (S - G_{3N})^2 r_i^2$  in column F. Compute the normalization constant  $N = \{4\pi \Delta r \sum_{i=1}^{300} [G_3(r_i)]^2 r_i^2\}^{-1/2}$  in some cell. Compute  $I$  in some cell. Graph  $S$  and  $G_{3N}$  on the same graph. Use the Solver in Excel to vary the six parameters in  $G_3$  to minimize  $I$ . Rerun the Solver with different initial guesses for the parameters. (c) For which range of  $r$  is  $G_{3N}$  significantly less than  $S$ ? Significantly greater than  $S$ ?

- 15.10** The fit in Prob. 15.9 gives  $S(r) \approx \sum_{i=1}^3 c_i G_i(r)$ , where  $S(r)$  is a 1s STO with orbital exponent 1 and  $G_i$  is a 1s GTO with orbital exponent  $\alpha_i$ . Let  $S(r, \zeta) \equiv \zeta^{3/2} \pi^{-1/2} e^{-\zeta r}$ . Let  $G_i(r, \zeta)$  be the function obtained by replacing  $\alpha_i$  in  $G_i(r)$  by  $\alpha_i \zeta^2$ . Show that  $S(r, \zeta) \approx \sum_{i=1}^3 c_i G_i(r, \zeta)$ .
- 15.11** Use the Basis Set Exchange ([bse.pnl.gov](http://bse.pnl.gov)) to find (a) the 6-31G\*\* basis functions for H; (b) the 6-31G\* basis functions for C.
- 15.12** Give an example of a molecule for which the HF/6-31G\* and HF/6-31G\*\* energies are the same.
- 15.13** For formaldehyde: (a) Work out the symmetry orbitals for a minimal-basis-set calculation; give the symmetry species of each symmetry orbital. (Choose the  $x$  axis perpendicular to the molecular plane.) (b) How many  $\sigma$  and how many  $\pi$  canonical MOs will result from a minimal-basis-set calculation? (See the Section 15.9 discussion of ethylene for the definition of  $\sigma$  and  $\pi$  MOs.) How many occupied  $\sigma$  and occupied  $\pi$  MOs are there for the ground state? (c) For each of the eight energy-localized MOs, state which AOs will make significant contributions. (d) What is the maximum-size secular determinant that occurs in finding the minimal-basis-set canonical MOs?
- 15.14** Work out the minimal-basis symmetry orbitals and their symmetry species for *cis*-1,2-difluoroethylene. Choose the  $x$  axis perpendicular to the molecular plane.
- 15.15** Sketch the  $1a_1, 3a_1, 1b_1, 4a_1$ , and  $2b_2$  MOs of water.
- 15.16** Give the form of the normalization constants for the symmetry orbitals in Eq. (15.13).
- 15.17** Suppose a ground-state calculation gives us some virtual orbitals for the molecule M. In which one of the following species would an excited electron occupy an MO that was well approximated by a virtual orbital of ground-state M? (a) M; (b)  $M^+$ ; (c)  $M^-$ . Explain.
- 15.18** Let  $E_{\text{HF}}$  be the Hartree–Fock energy of a closed-shell molecule; let  $E_{k,\text{HF,approx}}^+$  be the approximate Hartree–Fock energy of the ion formed by removal of an electron from the  $k$ th MO of this molecule, this energy being calculated using the MOs of the un-ionized molecule. For both the molecule and the ion, the Hartree–Fock wave function is a single determinant. Use Eqs. (11.80) to (11.82) [where (11.82) is modified to allow for the presence of several nuclei] and Eq. (14.30) to show that  $E_{\text{HF}} - E_{k,\text{HF,approx}}^+ = \epsilon_k$ , where  $\epsilon_k$  is the orbital energy of MO  $k$ .
- 15.19** Write symmetry orbitals for a minimal-basis-set calculation of  $\text{H}_2$ .
- 15.20** Suppose an SCF calculation that includes  $3d$  orbitals on oxygen is done on  $\text{H}_2\text{O}$  (Fig. 15.1). For each of the occupied ground-state  $\text{H}_2\text{O}$  MOs, use symmetry-species arguments to decide which of the following  $3d$  oxygen AOs will contribute to that MO:  $3d_{z^2}, 3d_{xz}, 3d_{yz}, 3d_{xy}, 3d_{x^2-y^2}$ .
- 15.21** (a) As noted in the text, for the cc-pVnZ basis sets, it is best not to use the  $n = 2$  energy value in extrapolations. Also, the  $n = 6$  energy value is often not available. Use (15.23) and the  $\text{H}_2\text{O}$  HF/cc-pVnZ energies for  $n = 3, 4$ , and 5 given just before (15.23) to find the values of the three parameters in (15.23). There are three equations in three unknowns.

Subtract the  $n = 3$  equation from the  $n = 4$  equation and from the  $n = 5$  equation to obtain two equations in two unknowns. Then combine these equations to eliminate  $A$ . To solve the remaining equation for  $B$ , put all terms on one side of the equation and use the Excel Solver to find  $B$ . Then find the other two unknowns. (b) Now use the  $n = 3, 4, 5$ , and  $6$  energies and the Excel Solver to find the values of the three parameters in (15.23) that minimize  $\sum_{n=3}^6 (E_{(15.23)} - E_{\text{cc-pVnZ}})^2$ . Take the initial guesses for the parameters as the results from part (a). In Excel 2010, be sure and uncheck the Make Unconstrained Variables Non-Negative box in the Solver Parameters box.

- 15.22** (a) Instead of (15.23), the following equation has been proposed for extrapolation from SCF aug-cc-pVQZ and aug-cc-pV5Z energies to the CBS limit [A. Karton and J. M. L. Martin, *Theor. Chem. Acc.*, **115**, 330 (2006)]:

$$E_{\text{SCF}}(\text{aug-}n) = E_{\text{SCF}}(\infty) + A(L + 1)e^{-9\sqrt{L}} \quad (15.88)$$

where  $L$  is the largest orbital-angular momentum  $l$  value that appears in the basis set. For  $\text{H}_2\text{O}$ , aug-cc-pVQZ and aug-cc-p5Z SCF energies at optimized geometries are  $-76.066676$  and  $-76.068009$  hartrees. Find the CBS limit predicted by this extrapolation formula. (b) Equation (15.88) can also be used with the aug-cc-pV5Z and aug-cc-pV6Z Hartree-Fock energies. The HF/aug-cc-pV6Z geometry-optimized energy for  $\text{H}_2\text{O}$  is  $-76.068153$  hartrees. Find the HF/CBS limit predicted by (15.88) and the aug-cc-pV5Z and aug-cc-pV6Z SCF energies. Compare the CBS values found in this problem to the estimate in Table 15.2.

- 15.23** (a) Use (15.24) to show that  $1 = \sum_r c_{ri}^2 + 2\sum_{r>s} \sum_s c_{ri} c_{si} S_{rs}$ . (b) Use the result of (a) to verify that  $\sum_r n_r + \sum_{r>s} \sum_s n_{r-s} = n$ .
- 15.24** Use the tables referred to in Prob. 15.29c [or one of the other available overlap-integral tables; see the reference given after (17.69)] to verify the values of the  $\text{H}_2\text{O}$  overlap integrals given in Section 15.6.
- 15.25** (a) Verify the net populations given at the end of the Section 15.6 example. (b) Verify the  $\text{H}_2\text{O}$   $2a_1$  and  $1b_2$  interatomic overlap populations given in Section 15.6. (c) Verify the gross populations given in Section 15.6 for the  $\text{H}_2\text{O}$  basis functions.
- 15.26** Use Coulomb's law to show that the electric potential  $\phi_P$  at a point P a distance  $d$  from a point charge  $Q$  is  $\phi_P = Q/4\pi\epsilon_0 d$ .
- 15.27** On Fig. 13.9 for ground-state  $\text{H}_2^+$ , sketch gradient paths that lie in the plane of the figure. Include the gradient path that does not end at a nucleus.
- 15.28** Frequently, the  $x$  and  $z$  axes of ethylene are interchanged as compared with Fig. 15.10. What change does this cause in the MO labeling? [The symmetry-species designations in Table 15.3 are retained.]
- 15.29** (a) Does having the coefficient of  $\text{C}1s$  greater than 1.0 in the  $1a_1$  methane MO in (15.46) violate the condition that the MO be normalized? Explain. (b) Use the results of Prob. 13.37 to express the  $2a_1$  methane MO of (15.46) using a nonorthogonal  $2s$  STO. (c) Verify that the  $2a_1$  MO found in part (b) is normalized. [The needed overlap integrals can be found by interpolating in the tables of R. S. Mulliken, C. A. Rieke, D. Orloff, and H. Orloff, *J. Chem. Phys.*, **17**, 1248 (1949).]
- 15.30** Call the ethylene symmetry orbitals in Table 15.4  $g_1$  to  $g_{14}$ , in order. For each of the eight canonical ethylene MOs occupied in the ground state, decide, as best you can, which symmetry orbitals will make major contributions; give the sign of each such symmetry orbital in the MO expression. (*Hint*: Decide how many inner-shell and how many bonding canonical MOs there are; combine the symmetry orbitals so as to build up charge density between the nuclei for the bonding MOs. *One further hint*: The third symmetry orbital in Table 15.4 makes no significant contribution to  $3a_g$ .) Sketch the canonical MOs. Assume an orthogonalized  $2s$  orbital is used.
- 15.31** Find the components of the gradient and the elements of the Hessian for each of the following functions, where the  $c$ 's are constants: (a)  $U = c_1 x^2 + c_2 y^2 + c_3 z^2$ ; (b)  $U = c(x + y + z)^2$ .

- 15.32** Give initial guesses for the geometry (bond lengths, bond angles, and dihedral angles) of the likely conformers of each of the following ground-state molecules. (a)  $\text{CH}_3\text{OH}$ ; (b)  $\text{C}_2\text{H}_4$ ; (c)  $\text{CH}_3\text{NH}_2$ ; (d)  $(\text{CH}_3)_2\text{CO}$ .
- 15.33** For the function  $U = 2x^2 - y^2$ , (a) find the stationary points; (b) find the first-order saddle points.
- 15.34** Consider the function  $U = 4(x - 1)^2 + 3(y - 2)^2$ . (a) By inspection, find its minimum. (b) Start at an arbitrary point  $x_1, y_1$  and show that the Newton–Raphson equation (15.72) with all partial derivatives evaluated exactly from  $U$  gives the minimum in one step.
- 15.35** For the function  $U = 3x^2 + 6y^2$ , (a) locate the minimum by inspection; (b) start at the point  $x = 9, y = 9$  and apply the conjugate-gradient method to locate the minimum. Do the calculations by hand. Recall that the equation of the straight line that passes through point  $x_1, y_1$  with slope  $m$  is  $m = (y - y_1)/(x - x_1)$ .
- 15.36** Write down the distance matrix for  $\text{H}_2\text{O}$  using the experimental geometry in Table 15.2. Take oxygen as atom 1.
- 15.37** True or false? (a) All triatomic molecules are planar. (b) The distance matrix is a symmetric matrix. (c) The diagonal elements of the distance matrix are zero. (d) The Hessian is a symmetric matrix.
- 15.38** As stated in Sections 15.3 and 15.5, the canonical MOs of a molecule can be chosen so that each one transforms according to one of the symmetry species of the molecule when the molecular symmetry operators are applied. The same is true of the normal-mode vibrations. For example, when each of the four symmetry operators of  $\text{H}_2\text{O}$  is applied to the  $\omega_1$  vibration in Fig. 15.16, the vibration displacement vectors end up looking unchanged. This normal mode therefore belongs to the totally symmetric symmetry species  $a_1$ . Give the symmetry species of the other two normal modes in Fig. 15.16. Take the axes as in Fig. 15.1.
- 15.39** The contribution of molecular vibrations to the molar internal energy  $U_m$  of a gas of nonlinear  $N$ -atom molecules is (zero-point vibrational energy not included)  $U_{m,\text{vib}} = R \sum_{s=1}^{3N-6} \theta_s / (e^{\theta_s/T} - 1)$ , where  $\theta_s \equiv h\nu_s/k$  and  $\nu_s$  is the vibrational frequency of normal mode  $s$ . Calculate the contribution to  $U_{m,\text{vib}}$  at  $25^\circ\text{C}$  of a normal mode with wavenumber  $\tilde{\nu} \equiv \nu_s/c$  of (a)  $900\text{ cm}^{-1}$ ; (b)  $300\text{ cm}^{-1}$ ; (c)  $2000\text{ cm}^{-1}$ .
- 15.40** (a) For  $\text{CO}_2$  a geometry-optimized HF/6-31G\* calculation gives  $U_e = -187.634176$  hartrees. UHF/6-31G\* ground-state energies are  $-37.680860$  and  $-74.783931$  hartrees for C and O, respectively. Calculate the predicted HF/6-31G\*  $D_e$  of  $\text{CO}_2$ . (b) Unscaled HF/6-31G\* vibrational wavenumbers for  $\text{CO}_2$  are  $745.8, 745.8, 1518.5$ , and  $2585.0\text{ cm}^{-1}$ . Scale these by  $0.89$ , find the HF/6-31G\* prediction for  $D_0$ , and compare with the experimental value  $D_0 = 16.56\text{ eV}$ . Also calculate the predicted atomization energy in kcal/mol. (c) Thermodynamic tables give  $\Delta H_{f,0}^\circ$  as  $246.79\text{ kJ/mol}$  for  $\text{O(g)}$  and  $711.2\text{ kJ/mol}$  for  $\text{C(g)}$  and give the molar enthalpy difference  $H_{m,298} - H_{m,0}$  as  $1.05\text{ kJ/mol}$  for graphite. Find the HF/6-31G\* predictions for  $\Delta H_{f,0}^\circ$  and  $\Delta H_{f,298}^\circ$  of  $\text{CO}_2(\text{g})$ . The experimental values are  $-393.2\text{ kJ/mol}$  and  $-393.5\text{ kJ/mol}$ .
- 15.41** Using standard bond lengths and bond angles, write Z-matrices using internal coordinates for each of the following: (a)  $\text{CO}_2$ ; (b)  $\text{CH}_4$ ; (c)  $\text{H}_2\text{CO}$ ; (d)  $\text{NH}_3$ ; (e)  $\text{C}_2\text{H}_4$ ; (f)  $\text{CH}_3\text{OH}$  (do both staggered and eclipsed conformations); (g)  $\text{CH}_2\text{ClCH}_2\text{Cl}$  (anti and gauche conformations).
- 15.42** Given the following Z-matrix (where the semicolons are used to denote the end of a row), draw a Newman projection of the molecule. C1; C2 1 1.5; F3 1 1.4 2 109.5; C14 2 1.8 1 120.0 3 0.0; H5 1 1.1 2 109.5 4 120.0; H6 1 1.1 2 109.5 4 -120.0; O7 2 1.2 1 120.0 3 180.0
- 15.43** If the dihedral angle  $D(\text{R}, \text{S}, \text{T}, \text{U})$  is  $60^\circ$ , what is the value of the dihedral angle  $D(\text{U}, \text{T}, \text{S}, \text{R})$ ?
- 15.44** For a nonlinear molecule containing  $N$  atoms, how many internal coordinates (bond lengths, bond angles, and dihedral angles) are specified in its Z-matrix? Justify your answer.
- 15.45** (a) For  $\text{CH}_2\text{O}$ , run HF/3-21G geometry optimization and vibrational-frequency calculations to obtain the predicted geometry, dipole moment, and harmonic vibrational wavenumbers. Verify that all vibrational wavenumbers are real. (b) Repeat (a) using a HF/6-31G\* calculation. Multiply the HF/6-31G\* harmonic wavenumbers by the  $0.895$  scale factor and compare

the results with the fundamental wavenumbers listed in the CCCBDB. (c) Using a program such as the WebMO Demo or Spartan Student that can show animations of each normal mode, view the modes and sketch each  $\text{CH}_2\text{O}$  normal mode using arrows to show atomic motions. For modes that involve out-of-plane vibrations, it is helpful to rotate the model as it vibrates. Also, have the program draw the predicted infrared spectrum. (d) From the predicted IR normal-mode intensities, which mode has the strongest absorption and which the weakest? (e) As noted in Prob. 15.38, each normal mode can be classified according to its symmetry species. Use the results of part (c) to give the symmetry species of each normal mode.

- 15.46** (a) Do an HF/6-31G\* calculation to find the predicted harmonic vibrational wavenumbers of  $\text{CH}_2=\text{CHCl}$  using a program that can show animations of vibrational modes. Examine the animations and answer the following questions. Give the wavenumber of the mode that shows the greatest amount of CC bond stretching. Give the wavenumber of the mode that shows the greatest amount of CCl bond stretching. Give the wavenumbers of the modes that involve out-of-plane motions. (To answer this, rotate the model so that the molecule is perpendicular to the plane of the screen.) Have the program draw the predicted IR spectrum. (b) Multiply the wavenumbers found in (a) by the scale factor 0.895 to give estimates of the fundamental frequencies and compare with the experimental fundamental frequencies, which can be found in the CCCBDB. (c) Use the wavenumbers found in (b) to estimate the zero-point energy of this molecule.
- 15.47** For this problem, use a program that can display molecular orbitals. Use the HF/6-31G\* method to display each of the occupied MOs and the lowest unoccupied MO (LUMO) of  $\text{H}_2\text{O}$ . Describe each occupied MO as predominantly inner-shell, bonding, or lone pair. For each lobe (region) of each MO, state which atoms are involved and whether the lobe is positive or negative. Give the HF/6-31G\* orbital energy and the symmetry species of each of these MOs. (If you are using the online WebMO Demo, after you get the optimized-geometry model using Gaussian, click on New Job Using This Geometry, click the right arrow, choose Gaussian, click the right arrow, then choose Molecular Orbitals as the Calculation. When you get the output model, scroll down to see the list of MOs and click the magnifying glass for the MO you want to view. You can right-click on an MO surface and choose Opacity to change the type of display. Be sure you have Jmol installed when using WebMO; [jmol.sourceforge.net](http://jmol.sourceforge.net).)
- 15.48** Repeat Prob. 15.47 for  $\text{C}_2\text{H}_4$ . Also state whether each MO is  $\sigma$  or  $\pi$ .
- 15.49** For this problem, use a program that can display molecular surface electrostatic potential (MSESP) maps. For each of the following molecules, use the HF/6-31G\* method to view the MSESP map at the HF/6-31G\* equilibrium geometry. State where the most negative regions and the most positive regions are located. (a) formamide; (b) ethene; (c) benzene; (d) cyclopropane. If you are using the online WebMO demo, follow the directions given in Prob. 15.47 to get the list of MOs. Then click on the magnifying glass opposite Electrostatic potential (which is below the MO list) to see the electrostatic potential map.
- 15.50** For  $\text{H}_2\text{O}$  use HF/6-31G\* to optimize the geometry and then calculate the ESP charges according to each of the methods Merz–Singh–Kollman, CHELP, and CHELPG. To do this in *Gaussian*, in addition to Opt on the line beginning with a #, one includes Pop=MK or Pop=CHELP, or Pop=CHELPG. If you are using the online WebMO Demo, after you get the optimized-geometry model using Gaussian, click on New Job Using This Geometry, click the right arrow, choose Gaussian, click the right arrow, then choose Other as the Calculation, and enter Pop=MK in the Enter Calculation window, click OK, and click the right arrow. Scroll down in the output model display to see the ESP charges. Compare with the Mulliken charges, which are listed as the Partial Charges by WebMO. If you have Spartan Student, after you optimize the geometry, go to the Model menu and choose Ball and Spoke; choose Configure on the Model menu, click Electrostatic Charge, click OK, and click Labels on the Model menu. Only one kind of ESP charge is given by Spartan Student.
- 15.51** The gas-phase benzene dimer species (weakly bound by dispersion forces) has been detected by microwave spectroscopy, and its dissociation energy  $D_0$  is about 2.4 kcal/mol. Two of the several possible structures for this dimer are the eclipsed  $\mathcal{D}_{6h}$  sandwich structure with the two

rings parallel to each other (like the two pieces of bread in a sandwich) and corresponding vertices of the two hexagons aligned, and a T structure with one hexagon above the other and perpendicular to it. Based on the ESP map of benzene (Prob. 15.49c), which is the more likely structure of these? [Actually, using the ESP map is much too crude a way to answer this question, and very high-level calculations are needed to determine the relative energies of benzene-dimer structures; see M. Pitoňák et al., *J. Chem. Theory Comput.*, **4**, 1829 (2008).] Aromatic  $\pi-\pi$  interactions are significant in proteins and DNA, so study of the benzene dimer has important applications.

- 15.52** Perform single-point HF/3-21G calculations on H<sub>2</sub>O with the bond lengths held constant at 0.96 Å and the bond angle taken as 100°, 102°, 104°, 106°, 108°, 110°, and 112°. Plot the energies versus bond angle and decide what order polynomial will give a good fit to the data. Use a graphing program or spreadsheet to fit the data with such a polynomial and find the 3-21G minimum-energy angle at this bond length.
- 15.53** In a Z-matrix for the *Gaussian* programs, letters can be used to specify some or all of the internal coordinates. For example, a Z-matrix for HOCl can be written as

```
O1
H2 1 R1
Cl3 1 R2 2 A1
Variables:
R1 0.96
R2 1.6
A1 108.0
```

If one wanted to hold the bond angle constant at 108° during a geometry optimization, the last line of the preceding Z-matrix would be replaced by

```
Constants:
A1 108.0
```

Do include a space before each of the words Variables and Constants. (In *Gaussian* 98 and 03, the keyword Opt must be replaced by Opt=Z-matrix.) Set up a Z-matrix for a partial geometry optimization for C<sub>2</sub>H<sub>6</sub> in which the dihedral angles are held fixed at values corresponding to (a) the staggered form; (b) the eclipsed form. Then perform HF/6-31G\*\* optimizations on these two conformations and take the energy difference to find the predicted barrier to internal rotation in kcal/mol.

- 15.54** Use CORINA on the Internet to build models and find Cartesian coordinates for the following molecules: (a) CH<sub>3</sub>OH; (b) HC(O)OH.
- 15.55** Use ChemDB to answer Prob. 15.54.
- 15.56** (a) Do a HF/6-31G\* geometry optimization and frequency calculation on NH<sub>3</sub> starting from a planar geometry. Is the final structure planar? Are all the vibrational frequencies real? (b) Repeat the calculation starting with a slightly nonplanar geometry. You might find it convenient to use a dummy atom in the Z-matrix. (c) What is the HF/6-31G\* predicted barrier to inversion of ammonia?
- 15.57** (a) Do HF/6-31G\* geometry optimizations on two conformers of HCOOH, one with OCOH dihedral angle of 0° and one with 180°. Compare the predicted geometries and the dipole moments of the two conformers. What is the predicted energy difference at 0 K omitting zero-point energy? (b) Do frequency calculations on these two conformers and use scaled vibrational frequencies to get the predicted 0 K energy difference including zero-point energy. Are all the vibrational frequencies real for each conformer?
- 15.58** (a) Do HF/6-31G\* partial geometry optimizations of *n*-butane conformations with CCCC dihedral angles fixed at several values. Plot the energy versus dihedral angle. From the plot, estimate the barriers for the conversions gauche → anti and anti → gauche. (b) Start from a 60° dihedral angle and do a full geometry optimization to find the predicted dihedral angle for the gauche conformer. Calculate the predicted energy difference between gauche and anti conformers at 0 K, neglecting zero-point energy. See Prob. 15.53.

- 15.59** Do HF/6-31G\* calculations to find the predicted geometries, dipole moments, and vibrational wavenumbers of the conformers of  $\text{N}_2\text{H}_4$ . What is the predicted energy difference (or differences) at 0 K?
- 15.60** Repeat Prob. 15.59 for propene,  $\text{CH}_2\text{CHCH}_3$ .
- 15.61** For a minimal-basis STO calculation on  $\text{H}_2$ , how many different electron-repulsion integrals ( $rs|tu$ ) need to be calculated, taking into account (14.47) and the symmetry of the molecule?
- 15.62** Slater's rules for finding approximate orbital exponents of  $K$ -,  $L$ -, and  $M$ -shell Slater AOs are as follows. The orbital exponent  $\zeta$  is taken as  $(Z - s_{nl})/n$ , where  $n$  is the principal quantum number and  $Z$  is the atomic number. The screening constant  $s_{nl}$  is calculated as follows: The AOs are divided into the following groups:

$$(1s) \quad (2s, 2p) \quad (3s, 3p) \quad (3d)$$

To find  $s_{nl}$ , the following contributions are summed: (a) 0 from electrons in groups to the right of the one being considered; (b) 0.35 from each other electron in the group considered, except that 0.30 is used in the  $1s$  group; (c) for an  $s$  or  $p$  orbital, 0.85 from each electron whose quantum number  $n$  is one less than the orbital considered and 1.00 from each electron still further in; for a  $d$  orbital, 1.00 for each electron in a group to the left.

Calculate the orbital exponents according to Slater's rules for the atoms H, He, C, N, O, S, and Ar. The optimum values of  $\zeta$  to use when approximating an AO as a single STO have been calculated and are given in E. Clementi and D. L. Raimondi, *J. Chem. Phys.*, **38**, 2686 (1963); E. Clementi et al., *J. Chem. Phys.*, **47**, 1300 (1967). Compare these optimum values with the above values found by Slater's rules. [For  $n = 4$ , Slater took  $\zeta$  as  $(Z - s_{nl})/3.7$ ; however, Slater's rules are generally unreliable for  $n \geq 4$ .]