

Semiempirical and Molecular-Mechanics Treatments of Molecules

Because of the difficulties in applying *ab initio* methods to medium and large molecules, many semiempirical methods have been developed to treat such molecules. The earliest semiempirical methods treated only the π electrons of conjugated molecules. This chapter begins with π -electron semiempirical methods (Sections 17.1 to 17.3) and then considers general semiempirical methods (Section 17.4).

The molecular-mechanics method (Section 17.5) is a nonquantum-mechanical method applicable to much larger molecules than semiempirical methods.

17.1 Semiempirical MO Treatments of Planar Conjugated Molecules

The canonical MOs of a planar unsaturated organic molecule can be divided into σ and π MOs according to whether the eigenvalue for reflection in the molecular plane is +1 or -1, respectively (Section 15.9). The earliest semiempirical methods for planar conjugated organic compounds (Sections 17.1–17.3) treated the π electrons separately from the σ electrons. Coulson stated that the justification for the σ – π separability approximation lies in the different symmetry of the σ and π orbitals and in the greater polarizability of the π electrons, which makes them more susceptible to perturbations such as those occurring in chemical reactions.

In the **π -electron approximation**, the n_π π electrons are treated separately by incorporating the effects of the σ electrons and the nuclei into some sort of effective π -electron Hamiltonian $\hat{H}_\pi$ (recall the similar valence-electron approximation; Sections 13.17 and 15.16):

$$\hat{H}_\pi = \sum_{i=1}^{n_\pi} \hat{H}_\pi^{\text{core}}(i) + \sum_{i=1}^{n_\pi} \sum_{j>i}^{n_\pi} \frac{1}{r_{ij}} \quad (17.1)$$

$$\hat{H}_\pi^{\text{core}}(i) = -\frac{1}{2} \nabla_i^2 + V(i) \quad (17.2)$$

where $V(i)$ is the potential energy of the i th π electron in the field produced by the nuclei and the σ electrons. The *core* is everything except the π electrons. The variational principle is then applied to find a π -electron wave function ψ_π that minimizes the variational integral

$\int \psi_\pi^* \hat{H}_\pi \psi_\pi d\tau$ to give a π -electron energy E_π . The validity of the π -electron approximation has been discussed by Lykos and Parr (see *Parr*, pages 41–45, 211–218). Since (17.1) is not the true molecular electronic Hamiltonian, treatments that make the π -electron approximation are semiempirical. The main π -electron MO theories are the Hückel MO method (Section 17.2) and the Pariser–Parr–Pople method (Section 17.3).

17.2 The Hückel MO Method

The most celebrated semiempirical π -electron theory is the **Hückel molecular-orbital (HMO) method**, developed in the 1930s to treat planar conjugated hydrocarbons. Here the π -electron Hamiltonian (17.1) is approximated by the simpler form

$$\hat{H}_\pi = \sum_{i=1}^{n_\pi} \hat{H}^{\text{eff}}(i) \quad (17.3)$$

where $\hat{H}^{\text{eff}}(i)$ somehow incorporates the effects of the π -electron repulsions in an average way. This sounds rather vague, and in fact the Hückel method does not specify any explicit form for $\hat{H}^{\text{eff}}(i)$. Since the Hückel π -electron Hamiltonian is the sum of one-electron Hamiltonians, a separation of variables is possible (Section 6.2). We have

$$\hat{H}_\pi \psi_\pi = E_\pi \psi_\pi \quad (17.4)$$

$$\psi_\pi = \prod_{i=1}^{n_\pi} \phi_i \quad (17.5)$$

$$\hat{H}^{\text{eff}}(i) \phi_i = e_i \phi_i \quad (17.6)$$

$$E_\pi = \sum_{i=1}^{n_\pi} e_i \quad (17.7)$$

where the product notation in (17.5) is defined by Eq. (17.22). The wave function (17.5) takes no account of spin or the antisymmetry requirement. To do so, we must put each electron in a spin-orbital $u_i = \phi_i \sigma_i$ where σ_i is a spin function (either α or β). The wave function ψ_π is then written as a Slater determinant of spin-orbitals. Since $\hat{H}^{\text{eff}}(i)$ does not involve spin, we have $\hat{H}^{\text{eff}}(i) u_i = e_i u_i$. Each term in the antisymmetrized product function ψ_π has each electron in a different spin-orbital [see, for example, the equation preceding (10.48)]. When $\hat{H}_\pi$, which is being approximated as the sum of the $\hat{H}^{\text{eff}}(i)$'s, acts on each term in ψ_π , it gives the sum of the e_i 's. Hence $\hat{H}_\pi \psi_\pi$ equals $\sum_i e_i \psi_\pi$, and (17.7) still holds when allowance is made for spin and antisymmetry.

Since $\hat{H}^{\text{eff}}$ is not specified, there is no point in trying to solve (17.6) directly. Instead, the variational method is used.

The next assumption in the HMO method is to approximate the π MOs as LCAOs. In a minimal-basis-set calculation of a planar conjugated hydrocarbon, the only AOs of π symmetry are the carbon $2p\pi$ orbitals, where by $2p\pi$ we mean the real $2p$ AOs that are perpendicular to the molecular plane. We thus write

$$\phi_i = \sum_{r=1}^{n_C} c_{ri} f_r \quad (17.8)$$

where f_r is a $2p\pi$ AO on the r th carbon atom and n_C is the number of carbon atoms. Since (17.8) is a linear variation function, the optimum values of the coefficients for the n_C lowest π MOs satisfy Eq. (8.53):

$$\sum_{s=1}^{n_C} [(H_{rs}^{\text{eff}} - S_{rs} e_i) c_{si}] = 0, \quad r = 1, 2, \dots, n_C \quad (17.9)$$

where the e_i 's are the roots of the secular equation (8.57):

$$\det(H_{rs}^{\text{eff}} - S_{rs}e_i) = 0 \quad (17.10)$$

The key assumptions in the Hückel theory involve the integrals in (17.10). The integral H_{rr}^{eff} is assumed to have the same value for every carbon atom in the molecule. (For benzene the six carbons are equivalent, and this is no assumption. For 1,3-butadiene, $\text{CH}_2\text{CHCHCH}_2$, one would expect H_{rr}^{eff} for an end carbon and a middle carbon to differ slightly.) Moreover, H_{rr}^{eff} is assumed to be the same for carbon atoms in different planar hydrocarbons. The integral H_{rs}^{eff} is assumed to have the same value for any two carbon atoms bonded to each other and to vanish for two nonbonded carbons. The integral S_{rr} is equal to 1, since the AOs are normalized. The overlap integral S_{rs} is assumed to vanish for $r \neq s$. We have

$$H_{rr}^{\text{eff}} = \int f_r^*(i) \hat{H}^{\text{eff}}(i) f_r(i) dv_i \equiv \alpha \quad (17.11)$$

$$H_{rs}^{\text{eff}} = \int f_r^*(i) \hat{H}^{\text{eff}}(i) f_s(i) dv_i \equiv \beta \quad \text{for } C_r \text{ and } C_s \text{ bonded} \quad (17.12)$$

$$H_{rs}^{\text{eff}} = 0 \quad \text{for } C_r \text{ and } C_s \text{ not bonded together} \quad (17.13)$$

$$S_{rs} = \int f_r^*(i) f_s(i) dv_i = \delta_{rs} \quad (17.14)$$

$$f_r \equiv C_r 2p\pi \quad (17.15)$$

where δ_{rs} is the Kronecker delta. α is called the **Coulomb integral** and β is the **bond integral** (or resonance integral). [Note that α is unrelated to the Coulomb integral in Eq. (14.24).] Since carbons not bonded to each other are well separated in space, the assumption (17.13) is reasonable. However, taking the overlap integral as zero for carbons bonded to each other is a poor assumption. For Slater orbitals, S_{rs} for adjacent carbons ranges between 0.2 and 0.3, depending on the bond distance. Inclusion of overlap will be considered later.

We want each π MO to be normalized. Using Eqs. (17.8) and (17.14), we have

$$1 = \langle \phi_i | \phi_i \rangle = \langle \sum_r c_{ri} f_r | \sum_s c_{si} f_s \rangle = \sum_r \sum_s c_{ri}^* c_{si} \langle f_r | f_s \rangle = \sum_r \sum_s c_{ri}^* c_{si} \delta_{rs} = \sum_r c_{ri}^* c_{ri}$$

$$\sum_{r=1}^{n_c} |c_{ri}|^2 = 1 \quad (17.16)$$

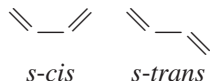
This is the normalization condition for the i th Hückel MO.

In application of HMO theory to conjugated hydrocarbons, planarity is assumed. Occasionally, this assumption does not hold. In gas-phase biphenyl the two rings are twisted at an angle of 44° with each other because of steric interference between the *ortho* hydrogens.

From (17.9) it is clear that the order of the HMO secular determinant equals the number of conjugated atoms. Students sometimes make the error of assuming that this order always equals the number of π electrons.

Butadiene

To illustrate the HMO method, we consider 1,3-butadiene. The only thing of significance in the simple HMO treatment of a planar hydrocarbon is the topology of the carbon-atom framework; by this we mean which carbons are bonded together. No distinction is made between *s-cis* and *s-trans* butadiene. The terms *s-cis* and *s-trans* refer to the conformation about the single bond:



The *s-trans* conformation is the dominant one observed for 1,3-butadiene. However, the UV, IR, and Raman spectra of 1,3-butadiene show the presence of small amounts of a second conformation. High-level CCSD(T) calculations and analysis of spectra of gas-phase butadiene show that this second conformation is the nonplanar gauche form with a CCCC dihedral angle of about 36° . The *cis* form is a saddle point that lies about 0.5 kcal/mol above the gauche minimum; the gauche conformer lies 3 kcal/mol above the trans global minimum. See D. Feller and N. C. Craig, *J. Phys. Chem. A*, **113**, 1601 (2009); P. Boopalachandran et al., *ibid.*, **115**, 8920 (2011); **116**, 271 (2012).

The numbering of the carbon atoms is



The Hückel assumptions (17.11) to (17.13) give $H_{11}^{\text{eff}} = H_{22}^{\text{eff}} = H_{33}^{\text{eff}} = H_{44}^{\text{eff}} = \alpha$, $H_{12}^{\text{eff}} = H_{23}^{\text{eff}} = H_{34}^{\text{eff}} = \beta$, and $H_{13}^{\text{eff}} = H_{14}^{\text{eff}} = H_{24}^{\text{eff}} = 0$. The secular equation (17.10) is

$$\begin{vmatrix} \alpha - e_k & \beta & 0 & 0 \\ \beta & \alpha - e_k & \beta & 0 \\ 0 & \beta & \alpha - e_k & \beta \\ 0 & 0 & \beta & \alpha - e_k \end{vmatrix} = 0 \quad (17.18)$$

We now divide each row of the determinant by β . This divides the determinant by β^4 , and since $0/\beta^4 = 0$, we get

$$\begin{vmatrix} x & 1 & 0 & 0 \\ 1 & x & 1 & 0 \\ 0 & 1 & x & 1 \\ 0 & 0 & 1 & x \end{vmatrix} = 0 \quad (17.19)$$

where

$$x \equiv \frac{\alpha - e_k}{\beta}, \quad e_k = \alpha - \beta x \quad (17.20)$$

A determinant in which all elements are zero except the elements of the principal diagonal and the elements immediately above and below this diagonal is called a *continuant*. When the elements immediately above the principal diagonal are all equal, those on the principal diagonal are all equal, and those immediately below the principal diagonal are all equal, the *n*th-order continuant can be shown to have the value (T. Muir, *The Theory of Determinants in the Historical Order of Development*, Dover, Volume 4, 1960, page 401)

$$\begin{vmatrix} a & b & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ c & a & b & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & c & a & b & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & c & a & b \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 0 & c & a \end{vmatrix} = \prod_{j=1}^n \left[a - 2(bc)^{1/2} \cos\left(\frac{j\pi}{n+1}\right) \right] \quad (17.21)$$

where n is the order of the determinant. The definition of the product notation used in (17.21) is

$$\prod_{j=1}^n b_j = b_1 b_2 b_3 \cdots b_n \quad (17.22)$$

Use of (17.21) in (17.19) gives

$$\prod_{j=1}^4 \left(x - 2 \cos \frac{j\pi}{5} \right) = 0$$

$$x = 2 \cos(j\pi/5), \quad j = 1, 2, 3, 4 \quad (17.23)$$

$$x = -1.618, -0.618, 0.618, 1.618 \quad (17.24)$$

Alternatively, we can expand the secular determinant to yield the algebraic equation $x^4 - 3x^2 + 1 = 0$. The substitution $z = x^2$ yields $z^2 - 3z + 1 = 0$, and the quadratic formula gives $z = (3 \pm 5^{1/2})/2$. Hence $x = \pm z^{1/2} = (3 \pm 5^{1/2})^{1/2}/2^{1/2}$, $-(3 \pm 5^{1/2})^{1/2}/2^{1/2}$.

Corresponding to (17.19), the equations for the HMO coefficients of butadiene are

$$\begin{aligned} xc_{1j} + c_{2j} &= 0 \\ c_{1j} + xc_{2j} + c_{3j} &= 0 \\ c_{2j} + xc_{3j} + c_{4j} &= 0 \\ c_{3j} + xc_{4j} &= 0 \end{aligned} \quad (17.25)$$

We must now substitute each of the roots (17.24) in turn into (17.25), as discussed in Section 8.5.

Consider first the root $x = -1.618$. The first equation of (17.25) reads $-1.618c_1 + c_2 = 0$ (where, for simplicity, the j subscript has been omitted). As discussed in Section 8.4, the solutions c_1, c_2, c_3, c_4 each contain an arbitrary multiplicative constant. Hence we shall solve for c_2, c_3 , and c_4 in terms of c_1 . We have $c_2 = 1.618c_1$. The second equation in (17.25) gives $c_3 = -c_1 - xc_2 = -c_1 + 1.618(1.618c_1) = 1.618c_1$. The fourth equation in (17.25) gives $c_4 = -c_3/x = 1.618c_1/1.618 = c_1$.

The normalization condition (17.16) is now used to fix c_1 . Taking c_1 to be a real, positive number, we have $1 = c_1^2 + c_2^2 + c_3^2 + c_4^2 = c_1^2 + (1.618c_1)^2 + (1.618c_1)^2 + c_1^2 = 7.236c_1^2$, and $c_1 = 0.372$. Then $c_2 = 1.618c_1 = 0.602$, $c_3 = 1.618c_1 = 0.602$, and $c_4 = c_1 = 0.372$. The HMO corresponding to $x = -1.618$ is then $\phi = 0.372f_1 + 0.602f_2 + 0.602f_3 + 0.372f_4$. The energy of this HMO is given by (17.20) as $e = \alpha + 1.618\beta$.

Substitution of each of the three remaining roots into (17.25) yields three more HMOs. We find the following four normalized HMOs (Prob. 17.4):

$$\begin{aligned} \phi_1 &= 0.372f_1 + 0.602f_2 + 0.602f_3 + 0.372f_4 \\ \phi_2 &= 0.602f_1 + 0.372f_2 - 0.372f_3 - 0.602f_4 \\ \phi_3 &= 0.602f_1 - 0.372f_2 - 0.372f_3 + 0.602f_4 \\ \phi_4 &= 0.372f_1 - 0.602f_2 + 0.602f_3 - 0.372f_4 \end{aligned} \quad (17.26)$$

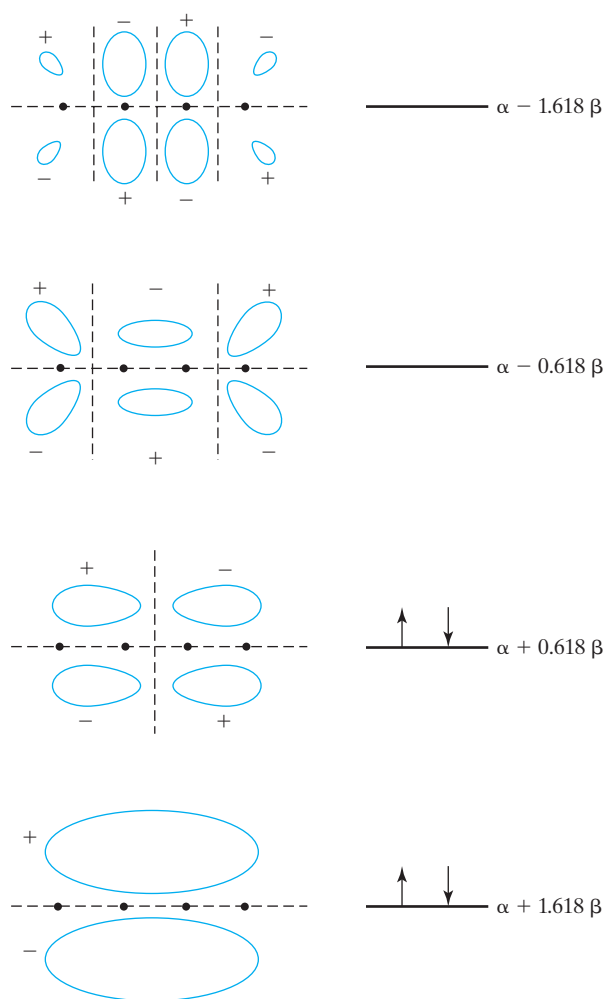


FIGURE 17.1 HMOs for butadiene. Nodal planes for the π MOs are indicated by dashed lines. The ground-state π -electron MO configuration is shown.

The HMO energies are

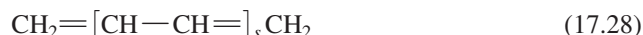
$$\begin{aligned} e_1 &= \alpha + 1.618\beta, & e_2 &= \alpha + 0.618\beta \\ e_3 &= \alpha - 0.618\beta, & e_4 &= \alpha - 1.618\beta \end{aligned} \quad (17.27)$$

The MO ϕ_1 in (17.26) has no nodes (other than the molecular plane) and leads to maximum charge buildup between the atoms. Clearly this must be the lowest-energy π MO. Its energy is $\alpha + 1.618\beta$ and, therefore, the bond integral β must be negative. [See also Eq. (13.62).] The HMOs and energy levels are sketched in Fig. 17.1. The number of vertical nodal planes is zero for the ground MO and increases by one for each higher MO.

From the figure, it is clear that these MOs are orthogonal. We can approximate the energy of an electron in a carbon $2p\pi$ AO in the molecule by $\int f_i^* \hat{H}^{\text{eff}} f_i dv = \alpha$. An HMO is classified as bonding or antibonding according to whether its energy is less than or greater than α .

Conjugated Polyenes

Consider the conjugated polyene



where s can be 0, 1, 2, ... Let n_C be the number of carbon atoms in the polyene. The HMO secular equation involves a continuant similar to (17.19) but of order n_C .

Equation (17.21) with $b = c = 1$, $a = x$, and $n = n_C$ gives $x = 2 \cos[j\pi/(n_C + 1)]$, where $j = 1, \dots, n_C$. Since the values of x for $j = k$ and for $j = n_C + 1 - k$ are simply the negatives of each other, we can write

$$x = -2 \cos \frac{j\pi}{n_C + 1}$$

$$e_j = \alpha + 2\beta \cos \frac{j\pi}{n_C + 1}, \quad j = 1, 2, \dots, n_C \quad (17.29)$$

Since β is negative, e_1 is the lowest π energy level. All the π -electron levels are nondegenerate. The HMO coefficients are (Prob. 17.5)

$$c_{rj} = \left(\frac{2}{n_C + 1} \right)^{1/2} \sin \frac{jr\pi}{n_C + 1} \quad (17.30)$$

In the ground electronic state of (17.28), the highest occupied and lowest vacant π MOs have $j = \frac{1}{2}n_C$ and $\frac{1}{2}n_C + 1$, respectively. HMO theory predicts the longest wavelength band of the electronic absorption spectrum of a conjugated polyene to occur at

$$\frac{1}{\lambda} = -\frac{4\beta}{hc} \sin \frac{\pi}{2n_C + 2} \quad (17.31)$$

where (17.29) and $\cos a - \cos b = -2 \sin[\frac{1}{2}(a + b)] \sin[\frac{1}{2}(a - b)]$ were used. The bond integral β is a semiempirical parameter and is adjusted to give the best fit to experimental data. If we use the observed butadiene longest-wavelength absorption $\lambda = 217$ nm, we find $|\beta|/hc = 37300 \text{ cm}^{-1}$, $|\beta| = 4.62 \text{ eV}$. With this value of β , we then calculate wavelengths for the polyenes (17.28). The predicted values do show the correct trend of increase in λ with increase in n_C , but agreement with experiment is poor. Predicted λ values show an average absolute error of 44% for the first several members of the series (Prob. 17.3). [The first excited HMO configuration gives rise to two electronic terms, a singlet and a triplet. In the HMO model, neglect of electronic repulsions gives the singlet and triplet terms the same energy. The observed longest wavelength electronic transition is a singlet–singlet transition, since singlet–triplet transitions are forbidden. Strictly speaking, we should use the average of the singlet–singlet and singlet–triplet energy differences to find β .]

An obvious defect of the Hückel approximation for conjugated polyenes is the use of a single value of β for each pair of adjacent carbon atoms. The bond lengths in these molecules alternate, and we expect β to be larger for doubly bonded carbons than for singly bonded carbons. Lennard-Jones used two polyene bond integrals β_1 and β_2 for C—C and C=C bonds, respectively [J. E. Lennard-Jones, *Proc. Roy. Soc.*, **A158**, 280 (1937)]. With $\beta_1 = -3.32 \text{ eV}$, $\beta_2 = -4.20 \text{ eV}$, agreement with experiment is much improved over the single- β predictions; the average absolute error of predicted λ values is reduced to 9%.

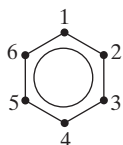


FIGURE 17.2 Numbering of carbons in benzene.

Benzene

For benzene (Fig. 17.2), the HMO secular equation is

$$\begin{vmatrix} x & 1 & 0 & 0 & 0 & 1 \\ 1 & x & 1 & 0 & 0 & 0 \\ 0 & 1 & x & 1 & 0 & 0 \\ 0 & 0 & 1 & x & 1 & 0 \\ 0 & 0 & 0 & 1 & x & 1 \\ 1 & 0 & 0 & 0 & 1 & x \end{vmatrix} = 0 \quad (17.32)$$

where x is given by (17.20).

The determinant in (17.32) is a special kind called a *circulant*. A circulant has only n independent elements; they appear in the first row, and succeeding rows are formed by successive cyclic permutations of these elements. The n th-order circulant $C(a_1, a_2, \dots, a_n)$ is

$$C(a_1, a_2, \dots, a_n) = \begin{vmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_n & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_n & a_1 & \cdots & a_{n-2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ a_2 & a_3 & a_4 & \cdots & a_1 \end{vmatrix}$$

The value of the n th-order circulant can be shown to be (T. Muir, *A Treatise on the Theory of Determinants*, Dover, 1960, pages 442–445)

$$C(a_1, a_2, \dots, a_n) = \prod_{k=1}^n (a_1 + \omega_k a_2 + \omega_k^2 a_3 + \cdots + \omega_k^{n-1} a_n) \quad (17.33)$$

where $\omega_1, \omega_2, \dots, \omega_n$ are the n different n th roots of unity [Eq. (1.36)]:

$$\omega_k = e^{2\pi i k/n}, \quad k = 1, 2, \dots, n, \quad i = \sqrt{-1} \quad (17.34)$$

Substitution of (17.33) into (17.32), followed by the use of $\exp[2\pi i k(5/6)] = \exp(-2\pi i k/6)$ and $e^{i\theta} = \cos \theta + i \sin \theta$, gives

$$\prod_{k=1}^6 (x + e^{2\pi i k/6} + e^{-2\pi i k/6}) = 0, \quad i = \sqrt{-1}$$

$$x = -2 \cos\left(\frac{2\pi k}{6}\right), \quad k = 1, \dots, 6 \quad (17.35)$$

$$x = -1, +1, +2, +1, -1, -2$$

$$e_i = \alpha + 2\beta, \quad \alpha + \beta, \quad \alpha + \beta, \quad \alpha - \beta, \quad \alpha - \beta, \quad \alpha - 2\beta \quad (17.36)$$

There are two nondegenerate and two doubly degenerate Hückel levels. **Figure 17.3** shows the HMO ground-state π -electron configuration.

The HMO coefficients can be found by solving the usual set of simultaneous equations, but it is simpler to use molecular symmetry. The $\hat{O}_{C_6}$ symmetry operator commutes with the π -electron Hamiltonian, so we can choose each MO to be an eigenfunction of this 60° rotation. Since $(\hat{O}_{C_6})^6 = \hat{I}$, the eigenvalues of $\hat{O}_{C_6}$ are the six sixth roots of unity (Prob. 7.25):

$$e^{2\pi i k/6}, \quad k = 0, 1, \dots, 5 \quad (17.37)$$

(Since $\hat{O}_{C_6}$ has some complex eigenvalues, it is not Hermitian. Only operators representing physical quantities need be Hermitian, and $\hat{O}_{C_6}$ does not correspond to any physical

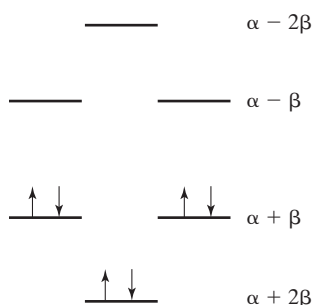


FIGURE 17.3 Hückel MO energies for benzene.

property of the molecule.) Substitution of $\phi_j = \sum_{r=1}^{n_c} c_{rj} f_r$ [Eq. (17.8)] into the eigenvalue equation $\hat{O}_{C_6} \phi_j = e^{2\pi i k/6} \phi_j$ gives

$$e^{2\pi i k/6} \phi_j = \hat{O}_{C_6} \phi_j$$

$$\sum_{r=1}^6 c_{rj} e^{2\pi i k/6} f_r = \sum_{r=1}^6 c_{rj} \hat{O}_{C_6} f_r = \sum_{r=1}^6 c_{rj} f_{r-1} = \sum_{r=1}^6 c_{r+1,j} f_r$$

where $f_0 \equiv f_6$ and $c_{7j} \equiv c_{1j}$. Equating the coefficients of corresponding AOs, we have

$$c_{r+1,j} = e^{2\pi i k/6} c_{rj} \quad (17.38)$$

The normalization condition (17.16) is $\sum_{r=1}^6 |c_{rj}|^2 = 1$ and Eq. (17.38) shows that all the coefficients in the j th MO have the same absolute value. Hence

$$|c_{rj}| = 1/\sqrt{6}, \quad r = 1, 2, \dots, 6 \quad (17.39)$$

By setting $c_{1j} = 1/\sqrt{6}$ and using (17.38), we get the desired coefficients. To find which coefficients go with which energy, we evaluate the variational integral:

$$e_j = \int \phi_j^* \hat{H}^{\text{eff}} \phi_j dv = \sum_{r=1}^6 \sum_{s=1}^6 c_{rj}^* c_{sj} \int f_r^* \hat{H}^{\text{eff}} f_s dv$$

$$= \sum_{r=1}^6 |c_{rj}|^2 \alpha + \sum_{r=1}^6 c_{rj}^* c_{r+1,j} \beta + \sum_{r=1}^6 c_{rj}^* c_{r-1,j} \beta$$

$$= \alpha + e^{2\pi i k/6} \sum_{r=1}^6 |c_{rj}|^2 \beta + e^{-2\pi i k/6} \sum_{r=1}^6 |c_{rj}|^2 \beta$$

$$= \alpha + 2\beta \cos(2\pi k/6), \quad k = 0, \dots, 5 \quad (17.40)$$

which agrees with (17.36). From (17.38), (17.39), and (17.40), the Hückel MOs and energies for benzene are

$$\begin{aligned} \phi_1 &= 6^{-1/2} (f_1 + f_2 + f_3 + f_4 + f_5 + f_6) \\ \phi_2 &= 6^{-1/2} (f_1 + e^{\pi i/3} f_2 + e^{2\pi i/3} f_3 - f_4 + e^{4\pi i/3} f_5 + e^{5\pi i/3} f_6) \\ \phi_3 &= 6^{-1/2} (f_1 + e^{-\pi i/3} f_2 + e^{-2\pi i/3} f_3 - f_4 + e^{-4\pi i/3} f_5 + e^{-5\pi i/3} f_6) \\ \phi_4 &= 6^{-1/2} (f_1 + e^{2\pi i/3} f_2 + e^{4\pi i/3} f_3 + f_4 + e^{2\pi i/3} f_5 + e^{4\pi i/3} f_6) \\ \phi_5 &= 6^{-1/2} (f_1 + e^{-2\pi i/3} f_2 + e^{-4\pi i/3} f_3 + f_4 + e^{-2\pi i/3} f_5 + e^{-4\pi i/3} f_6) \\ \phi_6 &= 6^{-1/2} (f_1 - f_2 + f_3 - f_4 + f_5 - f_6) \\ e_1 &= \alpha + 2\beta, \quad e_2 = \alpha + \beta, \quad e_3 = \alpha + \beta \\ e_4 &= \alpha - \beta, \quad e_5 = \alpha - \beta, \quad e_6 = \alpha - 2\beta \end{aligned} \quad (17.41)$$

The condition (17.38) determining the π -MO coefficients for benzene was derived solely from symmetry considerations, without use of the Hückel approximations. Thus *the MOs (17.41) are (except for normalization constants) the correct minimal-basis-set SCF π -electron MOs for benzene.* (The Hückel energies $e_1, \dots, e_6$ are, however, not the true SCF orbital energies. The Hückel method ignores electron repulsions and takes the total π -electron energy as the sum of orbital energies. The SCF MO method takes electron

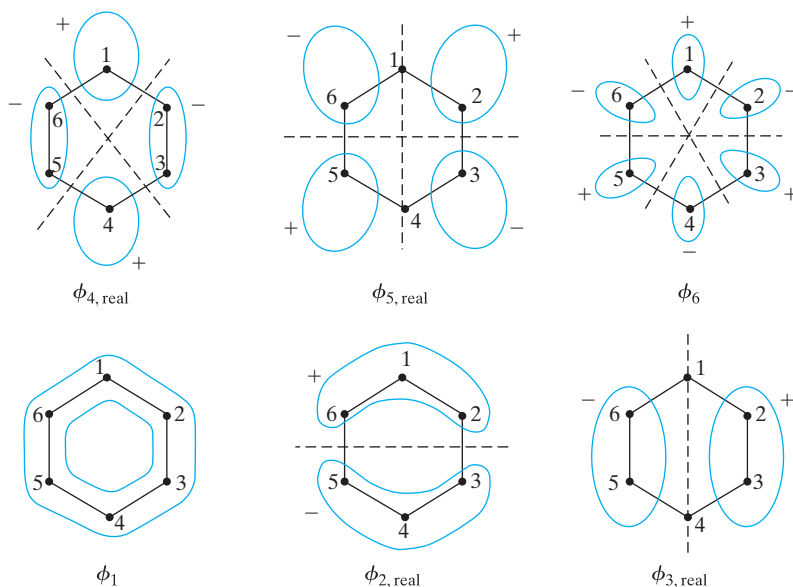


FIGURE 17.4 Benzene π MOs (real form). A top view is shown. The π MOs change sign on reflection in the molecular plane, which is a nodal plane for them. Dashed lines indicate nodal planes perpendicular to the molecular plane.

repulsions into account in an average way, and the total SCF energy is not the sum of orbital energies.) A similar situation occurs for ethylene, where the minimal-basis-set π MOs (Section 15.9) are determined solely by symmetry. (An extended-basis-set benzene calculation would mix in $3p\pi$, $3d\pi$, . . . , carbon AOs, and the contributions of these AOs must be determined by an explicit SCF calculation.)

The MOs in (17.41) for the degenerate π levels are complex. The two MOs of each degenerate level are complex conjugates of each other. By adding and subtracting these MOs, we get the commonly used real forms:

$$\begin{aligned}
 \phi_{2,\text{real}} &= 2^{-1/2}(\phi_2 + \phi_3), & \phi_{3,\text{real}} &= -2^{-1/2}i(\phi_2 - \phi_3) \\
 \phi_{2,\text{real}} &= 12^{-1/2}(2f_1 + f_2 - f_3 - 2f_4 - f_5 + f_6) \\
 \phi_{3,\text{real}} &= \frac{1}{2}(f_2 + f_3 - f_5 - f_6) \\
 \phi_{4,\text{real}} &= 12^{-1/2}(2f_1 - f_2 - f_3 + 2f_4 - f_5 - f_6) \\
 \phi_{5,\text{real}} &= \frac{1}{2}(f_2 - f_3 + f_5 - f_6)
 \end{aligned} \tag{17.42}$$

Figure 17.4 shows the real benzene π -electron MOs. Note the charge buildup between the nuclei for the bonding MOs.

The symmetry species of the benzene π MOs are (Schonland, p. 210)

MO	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
Symmetry species	a_{2u}	e_{1g}	e_{1g}	e_{2u}	e_{2u}	b_{2g}

The ground-state π -electron configuration is $(1a_{2u})^2(1e_{1g})^4$.

Monocyclic Conjugated Polyenes

For the monocyclic planar conjugated polyene C_nH_n , we can use the same treatment as for benzene, C_6H_6 . Replacing 6 by n_C in (17.40), (17.38), and (17.39), we find as the HMO energies and coefficients

$$e_k = \alpha + 2\beta \cos \frac{2\pi k}{n_C}, \quad k = 0, \dots, n_C - 1 \quad (17.43)$$

$$c_{rk} = \frac{1}{\sqrt{n_C}} \exp \left[\frac{2\pi i(r-1)k}{n_C} \right], \quad i = \sqrt{-1} \quad (17.44)$$

$$\phi_k = \frac{1}{\sqrt{n_C}} \sum_{r=1}^{n_C} \exp \left[\frac{2\pi i(r-1)k}{n_C} \right] f_r \quad (17.45)$$

where $f_r = C_r 2p\pi$. Note that the index k in these equations does not correspond to the actual order of the MOs: The lowest MO has $k = 0$; next are the MOs with $k = 1$ and $k = n_C - 1$; next are the MOs with $k = 2$ and $k = n_C - 2$; and so on.

An amusing mnemonic device is available for the HMO energies (17.43) [A.A. Frost and B. Musulin, *J. Chem. Phys.*, **21**, 572 (1953)]. One inscribes a regular polygon of n_C sides in a circle of radius $2|\beta|$, putting an apex at the bottom of the circle. If a vertical scale of energy is set up with energy α coinciding with the center of the circle, then each polygon vertex is located at an HMO energy (Fig. 17.5). The method gives the correct degeneracies and spacings of the Hückel levels of the ring hydrocarbon C_nH_n (Prob. 17.8).

Consider the Hückel energy levels for the monocyclic polyene C_nH_n . The lowest shell consists of a nondegenerate level and holds two electrons. Each of the remaining low-lying shells consists of a doubly degenerate level and holds four electrons. (If n_C is even, the highest π energy level is nondegenerate, but this level is not occupied in the ground state.) To have a stable filled-shell π -electron configuration, we see that the number of π electrons must satisfy

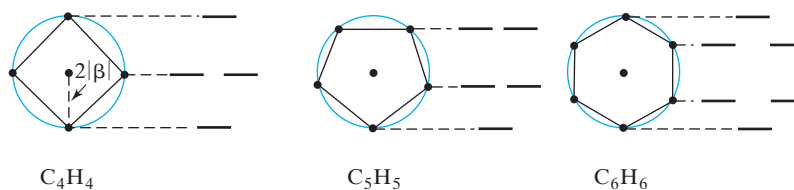
$$n_\pi = 4m + 2, \quad m = 0, 1, 2, \dots \quad (17.46)$$

This is Hückel's famous $4m + 2$ rule, which ascribes extra stability to monocyclic conjugated systems that satisfy (17.46). With $4m + 1$ or $4m - 1$ π electrons, the compound is a free radical. With $4m$ π electrons, there are two electrons in a shell that can hold four electrons, and Hund's rule predicts a triplet (diradical) ground state.

Benzene satisfies the $4m + 2$ rule. The cyclopentadienyl radical $\cdot C_5H_5$ is one electron short of satisfying (17.46); the cyclopentadienyl anion $C_5H_5^-$ satisfies the $4m + 2$ rule; the cation $C_5H_5^+$ is predicted to have a triplet ground state and be highly reactive. These predictions are borne out; $C_5H_5^-$ is found to be considerably more stable than either $C_5H_5^+$ or $\cdot C_5H_5$. Similarly, $C_7H_7^+$ should be more stable than $\cdot C_7H_7$ or $C_7H_7^-$, as is verified experimentally; for example, the salt $C_7H_7^+Br^-$ is readily prepared.

C_4H_4 is predicted by the $4m + 2$ rule to have a triplet ground state. Cyclobutadiene, C_4H_4 , first synthesized in 1965, is a highly reactive compound that dimerizes at temperatures

FIGURE 17.5 Mnemonic device for HMO energies of monocyclic planar hydrocarbons. The levels shown are for cyclobutadiene, the cyclopentadienyl radical, and benzene.



above 35 K. Experimental observations and theoretical calculations agree that the ground electronic state is a singlet with a rectangular geometry and carbon–carbon bond lengths close to ordinary single- and double-bond lengths [see D. W. Whitman and B. K. Carpenter, *J. Am. Chem. Soc.*, **102**, 4272 (1980); W. T. Borden and E. R. Davidson, *Acc. Chem. Res.*, **14**, 69 (1981)]. The reason for this violation of Hund’s rule is discussed in H. Kollmar and V. Staemmler, *J. Am. Chem. Soc.*, **99**, 3583 (1977).

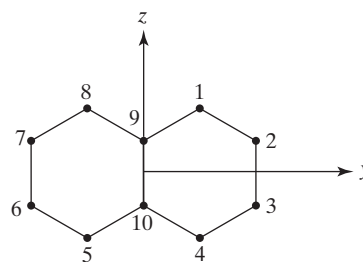
HMO theory predicts that planar cyclooctatetraene, C_8H_8 , would have a triplet ground state. Experimentally, C_8H_8 has a nonplanar “tub” structure with alternating single and double bonds. The nonplanarity results from steric strain in the planar geometry due to deviation from the 120° bond angle at the sp^2 -hybridized carbons. Therefore, C_8H_8 does not provide a good test of the $4m + 2$ rule. The dianion $C_8H_8^{2-}$ satisfies the $4m + 2$ rule. The salt $K_2C_8H_8$ has been prepared, and the evidence is that $C_8H_8^{2-}$ is planar. The extra stability arising from delocalization of the π electrons is sufficient to overcome the steric strain.

The monocyclic compounds C_nH_n are called *annulenes*. Benzene is [6]annulene. The annulenes with $n = 10, 12, 14$, and 16 suffer steric strain and substantial repulsions between nonbonded hydrogens, thereby preventing planarity and a clear-cut test of the $4m + 2$ rule. The compound [18]annulene has been much studied, but the results are conflicting. [18]annulene has a nearly planar 120° -angle structure with a cis configuration about every third double bond. This makes 6 of the 18 hydrogens lie in the interior of the ring. X-ray crystallography data of the solid have been interpreted as showing nearly equal carbon–carbon bond lengths, but this result is not conclusive. Various high-level ab initio calculations give differing results as to whether the bond lengths alternate. Theoretical calculations of the NMR proton shifts in [18]annulene found that good agreement with the experimental values was achieved only if the molecular geometry showed a carbon–carbon bond-length alternation of about $0.1 \text{ \AA}$ [C. S. Wannere et al., *Angew. Chem. Int. Ed.*, **43**, 4200 (2004)]. DFT calculations in this paper found that the $\mathcal{D}_{6h}$ structure with nearly equal bond lengths was not an energy minimum, and found a $\mathcal{C}_2$ structure with substantial bond-length alternation to be the energy minimum. That the $\mathcal{C}_2$ structure was lower in energy than the $\mathcal{D}_{6h}$ structure was confirmed by a single-point CCSD(T) calculation with a moderate-size basis set. The $\mathcal{D}_{6h}$ structure was found to be a transition state between bond-alternating $\mathcal{C}_2$ structures. The authors argued that since the $\mathcal{D}_{6h}$ structure is only 2 or 3 kcal/mol higher in energy than equivalent equilibrium $\mathcal{C}_2$ structures, rapid interconversion makes the X-ray data seem to show nearly equal bond lengths. Despite this bond-length alternation, the authors of this paper view [18]annulene as an aromatic compound because of its considerable aromatic stabilization energy and the abnormal values of its proton NMR chemical shifts.

In the preceding discussion, we used the same HMOs for a neutral compound and for its related ions. The Hückel method ignores electron repulsions, so the HMOs are unchanged when π electrons are added or removed.

The $4m + 2$ rule actually does not depend on the Hückel assumptions (17.11) to (17.14). The C_nH_n π MOs (17.45) were derived solely by symmetry considerations and are the correct SCF minimal-basis-set π MOs. With $k = 0$, we have an MO with all plus signs in front of the AOs. Clearly this MO has a lower energy than any of the others. For the remaining MOs, the pair with $k = j$ and $k = n_C - j$ are complex conjugates of each other and must have the same energy. [Since $\hat{H}$ is Hermitian, we have $\int \phi^* \hat{H} \phi dv = \int (\phi^*)^* \hat{H} \phi^* dv$.] Thus the excited MOs occur in pairs (except that when n_C is even, the MO with $k = \frac{1}{2}n_C$ is nondegenerate; with alternating plus and minus signs in front of the AOs, this is the highest-energy MO). The pattern of a nondegenerate lowest π level followed by doubly degenerate π levels thus holds for the minimal-basis-set SCF MOs. Of course, the energies (17.43) are not the correct SCF orbital energies.

FIGURE 17.6 Axes for naphthalene. The x axis is perpendicular to the molecular plane.



Naphthalene

Now consider the HMO treatment of naphthalene. For butadiene and benzene, we set up the secular equation without bothering with the intermediate step of constructing symmetry orbitals from the $2p\pi$ AOs. For these molecules, the secular equation was easy enough to solve without the simplifications introduced by symmetry orbitals. For naphthalene the 10×10 secular determinant is difficult to deal with, and we first find symmetry orbitals. The point group of naphthalene (Fig. 17.6) is $\mathcal{D}_{2h}$.

The possible symmetry species are in Table 15.3. The $C_{2p\pi}$ AOs all have eigenvalue -1 for reflection in the molecular (yz) plane. Each symmetry orbital will be some linear combination of AOs that are permuted among one another by the symmetry reflections (recall ethylene). Hence, to aid in finding the symmetry orbitals, we examine the effects of the $\hat{\sigma}(xy)$ and $\hat{\sigma}(xz)$ operations on the π AOs. We find the π AOs to fall into three sets:

$$1, 4, 5, 8 \quad 2, 3, 6, 7 \quad 9, 10$$

where the members of each set are permuted among one another by the symmetry reflections. (These are also the chemically equivalent carbons.) Each symmetry orbital must be some linear combination of the AOs of a given set. Naphthalene and ethylene have the same point group, and the pattern of the hydrogen symmetry orbitals in Table 15.4 gives us the symmetry orbitals of the first two of the above sets of naphthalene AOs. The symmetry orbitals and their readily verified symmetry species are then

$$\begin{aligned} b_{3u}: \quad g_1 &= \tfrac{1}{2}(f_1 + f_4 + f_5 + f_8), \quad g_2 = \tfrac{1}{2}(f_2 + f_3 + f_6 + f_7), \quad g_3 = 2^{-1/2}(f_9 + f_{10}) \\ a_u: \quad g_4 &= \tfrac{1}{2}(f_1 - f_4 + f_5 - f_8), \quad g_5 = \tfrac{1}{2}(f_2 - f_3 + f_6 - f_7) \\ b_{2g}: \quad g_6 &= \tfrac{1}{2}(f_1 - f_4 - f_5 + f_8), \quad g_7 = \tfrac{1}{2}(f_2 - f_3 - f_6 + f_7), \quad g_8 = 2^{-1/2}(f_9 - f_{10}) \\ b_{1g}: \quad g_9 &= \tfrac{1}{2}(f_1 + f_4 - f_5 - f_8), \quad g_{10} = \tfrac{1}{2}(f_2 + f_3 - f_6 - f_7) \end{aligned}$$

The constants $\frac{1}{2}$ and $1/\sqrt{2}$ normalize the symmetry orbitals, provided the approximation $S_{rs} = \delta_{rs}$ [Eq. (17.14)] is used.

Instead of using the f_r 's as basis functions, we set up the secular equation using the symmetry orbitals as basis functions:

$$\det[\langle g_p | \hat{H}^{\text{eff}} | g_q \rangle - \langle g_p | g_q \rangle e_k] = 0$$

The secular determinant is in block-diagonal form (Section 15.5), and we have two cubic and two quadratic equations to solve. Using $S_{rs} = \delta_{rs}$, we find that $\langle g_p | g_q \rangle = \delta_{pq}$ when g_p and g_q belong to the same symmetry species. Evaluating the matrix elements, we get as the secular equation for the b_{3u} MOs (Prob. 17.15)

$$\begin{vmatrix} \alpha - e_k & \beta & \sqrt{2}\beta \\ \beta & \alpha + \beta - e_k & 0 \\ \sqrt{2}\beta & 0 & \alpha + \beta - e_k \end{vmatrix} = 0, \quad \begin{vmatrix} x & 1 & \sqrt{2} \\ 1 & x+1 & 0 \\ \sqrt{2} & 0 & x+1 \end{vmatrix} = 0$$

$$(x+1)(x^2+x-3)=0$$

$$x = -1, \quad -\frac{1}{2} \pm \frac{1}{2}\sqrt{13}$$

Solution of the three remaining secular equations is left as an exercise. The naphthalene HMO energy levels in order of increasing energy are found to be (Prob. 17.15)

$$\alpha + 2.303\beta, \quad \alpha + 1.618\beta, \quad \alpha + 1.303\beta, \quad \alpha + \beta, \quad \alpha + 0.618\beta$$

$$\alpha - 0.618\beta, \quad \alpha - \beta, \quad \alpha - 1.303\beta, \quad \alpha - 1.618\beta, \quad \alpha - 2.303\beta$$

The levels are all nondegenerate ($\mathcal{D}_{2h}$ has only *A* and *B* symmetry species).

The HMO coefficients are found by solving the appropriate sets of simultaneous equations (Prob. 17.15).

Alternant Hydrocarbons

Note that the HMOs of naphthalene, like those of butadiene and benzene, are paired, meaning that for each HMO with the energy $\alpha - x\beta$ there is an HMO with energy $\alpha + x\beta$. This can be proved to be true for every alternant hydrocarbon (for the proof, see Prob. 17.21). An **alternant hydrocarbon** is a planar conjugated hydrocarbon in which the carbon atoms can be divided into a starred set and an unstarred set, with starred carbons bonded only to unstarred carbons, and vice versa. All planar conjugated hydrocarbons are alternants except those containing a ring with an odd number of carbons.

Electronic Transitions

The predicted wavenumber for a transition between the highest occupied and lowest vacant HMOs of a conjugated hydrocarbon is

$$\frac{1}{\lambda} = \frac{|\beta|}{hc} \Delta x \quad (17.47)$$

where Δx is the difference in x values [Eq. (17.20)] for the two MOs. For naphthalene, $\Delta x = 1.236$, and the observed $1/\lambda$ is 34700 cm^{-1} . Choosing β to fit the observed wavelength for naphthalene (because of the orbital degeneracy of its first excited term, benzene is atypical), we find

$$|\beta|/hc = 28100 \text{ cm}^{-1}, \quad |\beta| = 3.48 \text{ eV}$$

Comparison of predicted and observed longest-wavelength absorptions for benzenoid hydrocarbons shows only fair agreement with experiment, with many deviations of a few thousand cm^{-1} .

Improved agreement can be obtained if we fit the frequencies to a straight line that does not pass through the origin; that is, we use

$$\frac{1}{\lambda} = \frac{1}{hc} |\beta| \Delta x + a \quad (17.48)$$

A least-squares fit of benzenoid hydrocarbon data gives [E. Heilbronner and J. N. Murrell, *J. Chem. Soc.*, **1962**, 2611]

$$a = 8200 \text{ cm}^{-1}, \quad |\beta|/hc = 21900 \text{ cm}^{-1}, \quad |\beta| = 2.72 \text{ eV} \quad (17.49)$$

These constants give a good fit to the data; the standard deviation is 600 cm^{-1} . Of course, with two parameters instead of one, the agreement is bound to be improved. The fact that a semiempirical theory with several adjustable parameters gives a good fit to experimental data cannot be taken as overwhelming proof of the validity of the theory.

We can partially justify the constant term in (17.48) as follows. The Hückel method neglects electron repulsions and therefore does not distinguish between singlet and triplet

terms. Hence we should compare (17.47) with the energy difference between the ground state and the average of the energies of the singlet and triplet terms of the configuration with one electron excited to the lowest vacant HMO. The experimental frequencies are for singlet–singlet transitions. If we assume that the singlet–triplet splitting of the first excited configuration is reasonably constant for aromatic hydrocarbons, then $a = 8200 \text{ cm}^{-1} = 1.0 \text{ eV}$ can be interpreted as one-half this singlet–triplet splitting. Experimental values for half the singlet–triplet splitting in aromatic hydrocarbons are typically 0.7 to 0.8 eV, in reasonable agreement with $a = 1.0 \text{ eV}$.

Delocalization Energy and Aromaticity

Several properties of conjugated hydrocarbons can be defined using HMO theory. We begin with **delocalization energy**. The corresponding VB term is *resonance energy*. The energy of the occupied ethylene π HMO, the $1b_{3u}$ MO in Section 15.9, is

$$\int \frac{1}{\sqrt{2}}(f_1 + f_2)^* \hat{H}^{\text{eff}} \frac{1}{\sqrt{2}}(f_1 + f_2) dv = \alpha + \beta$$

There are two π electrons in this MO, and the total Hückel π -electron energy for ethylene is $2\alpha + 2\beta$. For butadiene the total Hückel π -electron energy is

$$2(\alpha + 1.618\beta) + 2(\alpha + 0.618\beta) = 4\alpha + 4.472\beta \quad (17.50)$$

If butadiene had two isolated double bonds, its π -electron energy would be twice that of ethylene, namely, $4\alpha + 4\beta$. The effect of delocalization is to change the butadiene π -electron energy by

$$4\alpha + 4.472\beta - (4\alpha + 4\beta) = 0.472\beta$$

Since β is negative, butadiene is stabilized by π -electron delocalization, and $0.472 |\beta|$ is its HMO delocalization energy.

For benzene the π -electron energy is

$$2(\alpha + 2\beta) + 4(\alpha + \beta) = 6\alpha + 8\beta \quad (17.51)$$

as compared with $6\alpha + 6\beta$ for the π -electron energy of three isolated double bonds. The delocalization energy of benzene is $2 |\beta|$. An “experimental” delocalization energy for benzene can be calculated as follows. The gas-phase enthalpy of hydrogenation of cyclohexene to cyclohexane is -28.6 kcal/mol . If benzene had three isolated double bonds, its gas-phase enthalpy of hydrogenation to cyclohexane would be three times -28.6 kcal/mol , which is -85.8 kcal/mol . The observed value is only -49.8 kcal/mol , indicating that benzene is more stable by 36 kcal/mol , than it would be if its double bonds were isolated. (A similar delocalization energy is arrived at by adding up the bond energies of six C—H bonds, three C—C bonds, and three C=C bonds and comparing this to minus the enthalpy of formation of gas-phase benzene from its atoms; see Prob. 17.10.) Setting $2 |\beta| = 36 \text{ kcal/mol}$, we get $|\beta| = 18 \text{ kcal/mol} = 0.8 \text{ eV/molecule}$.

This $|\beta|$ is far less than the value $2.72 \text{ eV/molecule} = 63 \text{ kcal/mol}$ found by spectroscopic observations [Eq. (17.49)]. Part of the discrepancy is explainable as follows. In the hydrogenation of cyclohexene, the carbon–carbon double-bond length becomes a single-bond length. The figure of -85.8 kcal/mol applies to the hydrogenation of a hypothetical molecule with three isolated double bonds and three single bonds, with alternating bond lengths. We must therefore also consider the strain energy needed to compress three single bonds and stretch three double bonds to the benzene bond length. Even with this correction (Prob. 17.9), the thermochemical value of β differs sharply from the spectroscopic value. The difference is to be attributed to the crudity of the

HMO method. One finds that a different value of β is required for each different physical property that is being considered. Moreover, the optimum values of β differ for chain and ring conjugated hydrocarbons.

The just-discussed traditional method of calculating Hückel delocalization energies of conjugated hydrocarbons by comparing a molecule's HMO π -electron energy E_π with $n_d(2\alpha + 2\beta)$, where n_d is the number of carbon-carbon double bonds, has serious shortcomings. This method predicts a substantial delocalization energy for the linear polyenes (17.28), whereas experiment shows the delocalization stabilization in these molecules to be small. For example, comparison of twice the enthalpy of hydrogenation of 1-butene with the enthalpy of hydrogenation of 1,3-butadiene gives the 1,3-butadiene delocalization energy as only 4 kcal/mol. Moreover, this method predicts substantial delocalization stabilization for certain cyclic polyenes that experiment reveals to be unstable, with no aromatic character. (A cyclic conjugated polyene is said to be **aromatic** when it shows substantially more stability than a hypothetical structure in which the double bonds do not interact with one another and when it undergoes substitution, rather than addition, when treated with electrophilic reagents like Br_2 .)

To produce more reliable predictions of aromaticity, Hess and Schaad (following a suggestion of Dewar) calculated delocalization (resonance) energies of cyclic hydrocarbons by comparing the compounds' Hückel theory E_π with a value calculated for a hypothetical acyclic conjugated polyene with the same number and kinds of bonds as in a localized structure of the cyclic hydrocarbon. [B. A. Hess and L. J. Schaad, *J. Am. Chem. Soc.*, **93**, 305, 2413 (1971); **94**, 3068 (1972); **95**, 3907 (1973); Schaad and Hess, *J. Chem. Educ.*, **51**, 640 (1974); Hess and Schaad, *Pure Appl. Chem.*, **52**, 1471 (1980); Schaad and Hess, *Chem. Rev.*, **101**, 1465 (2001).]

These workers found that the Hückel π -electron energies of noncyclic conjugated polyenes could be accurately calculated as $E_\pi \approx \sum_b n_b E_{\pi,b}$, where n_b is the number of bonds of a given type, $E_{\pi,b}$ is an empirical parameter, and the sum goes over all the types of carbon-carbon bonds. A conjugated hydrocarbon has three types of carbon-carbon single bonds and five types of double bonds, the types differing in the number of H atoms bonded to the carbons. Having found values of the parameters $E_{\pi,b}$ by fitting Hückel energies of noncyclic conjugated polyenes, Hess and Schaad then calculated the resonance energy of a cyclic polyene as the difference $\sum_b n_b E_{\pi,b} - E_\pi$ between the sum $\sum_b n_b E_{\pi,b}$ for a localized structure of the cyclic polyene and the cyclic polyene's Hückel π -electron energy E_π . For details, see Prob. 17.13.

Division of the Hess-Schaad resonance energy by the number of π electrons gives the resonance energy per π electron (REPE) of the compound. A compound with a substantially positive REPE value (greater than, say, $0.01 |\beta|$) is predicted to be *aromatic*. A compound with a near-zero REPE is *nonaromatic*. A compound with a substantially negative REPE is predicted to be *antiaromatic*, being less stable than if its double bonds were isolated from one another. Some antiaromatic hydrocarbons are cyclobutadiene and fulvalene.

For the annulenes, C_nH_n , Hess-Schaad REPE values are positive for $n = 4m + 2$ and negative for $n = 4m$. As n increases, the REPE becomes negligible, and the aromatic $4m + 2$ compounds and antiaromatic $4m$ compounds become nonaromatic. For example, $\text{REPE}/|\beta|$ values for $n = 4, 6, 8$ are $-0.268, 0.065$, and -0.061 , respectively, and $\text{REPE}/|\beta|$ values for $n = 16, 18$, and 20 are $-0.011, 0.012$, and -0.005 , respectively.

Comparison with experiment shows that the Hess-Schaad method is quite successful in predicting aromaticity.

For further discussion of the controversial concept of aromaticity, see V. I. Minkin et al., *Aromaticity and Antiaromaticity*, Wiley, 1994; *Aromaticity and Other Conjugation Effects*, R. Gleiter and G. Haberhauer, Wiley-VCH, 2012. The May 2001 and October 2005 issues of *Chemical Reviews* are devoted to aromaticity and related topics.

π -Electron Charges and Bond Orders

Another defined quantity is the π -electron charge. The probability density for an electron in the HMO (17.8) is

$$|\phi_i|^2 = \sum_r \sum_s c_{ri}^* c_{si} f_r^* f_s \quad (17.52)$$

The HMO normalization condition (17.16) is $\sum_{r=1}^{n_c} |c_{ri}|^2 = 1$, and so it is natural to say that an electron in the MO ϕ_i has the probability $|c_{ri}|^2$ of being in the vicinity of the r th carbon atom. If n_i ($= 0, 1$, or 2) is the number of electrons in the MO ϕ_i , then the total **π -electron charge** q_r in the region of carbon-atom r is defined as

$$q_r \equiv \sum_i n_i |c_{ri}|^2 \quad (17.53)$$

where the sum is over the π MOs. q_r is often called the π -electron density; this name is misleading, since q_r is neither a charge density (which has dimensions of charge/volume) nor a probability density (which has dimensions of 1/volume). Rather, q_r is a pure number that gives the approximate number of π electrons in the vicinity of carbon atom r .

For 1,3-butadiene, we have $q_1 = 2|c_{11}|^2 + 2|c_{12}|^2 = 2(0.372)^2 + 2(0.602)^2 = 1.000$ and $q_2 = 2|c_{21}|^2 + 2|c_{22}|^2 = 1.000$.

For the ground state of a neutral alternant hydrocarbon, all the Hückel π -electron charges q_r are 1.

In the population-analysis discussion of Section 15.6, we saw that for a real MO ϕ_i expanded as the linear combination $\sum_r c_{ri} f_r$ of AOs f_r , the quantity $2c_{ri}c_{si}S_{rs}$ in Eq. (15.25) (where S_{rs} is the overlap integral) is a reasonable measure of the contribution of an electron in MO ϕ_i to the bonding overlap between AOs f_r and f_s . In the HMO method, overlap integrals are neglected, so we cannot use this population-analysis expression. The carbon-carbon bond distances in a conjugated compound are all reasonably similar, so we expect the $C2p\pi$ – $C2p\pi$ overlap integral S_{rs} to have similar values for all pairs of bonded carbons. For nonbonded carbons, S_{rs} should be quite small. Since S_{rs} is approximately constant for bonded atoms, we can ignore the factor S_{rs} (and the constant factor 2) and take $c_{ri}c_{si}$ as the contribution of an electron in the real HMO ϕ_i to the π -electron bonding between bonded atoms r and s . If the MO ϕ_i is complex rather than real, one finds on integrating $|\phi_i|^2$ to produce the equation corresponding to (15.24) that the quantity $\frac{1}{2}(c_{ri}^*c_{si} + c_{si}^*c_{ri})$ occurs instead of $c_{ri}c_{si}$. Coulson therefore defined the **π -electron (or mobile) bond order** p_{rs} for the bond between bonded atoms r and s as

$$p_{rs} \equiv \frac{1}{2} \sum_i n_i (c_{ri}^* c_{si} + c_{si}^* c_{ri}) \quad (17.54)$$

where the sum is over the π MOs. When the coefficients are all real, (17.54) reduces to $p_{rs} = \sum_i n_i c_{ri}c_{si}$. That this definition is reasonable is indicated by the fact that it gives $p = 1$ for the π bond in ethylene.

Addition of the σ electrons' single bond gives the *total bond order* p_{rs}^{tot} as

$$p_{rs}^{\text{tot}} = 1 + p_{rs} \quad (17.55)$$

For butadiene we have $p_{12} = 2(0.372)(0.602) + 2(0.602)(0.372) = 0.894$ and $p_{23} = 2(0.602)(0.602) + 2(0.372)(-0.372) = 0.447$. The total bond orders are

$$\text{CH}_2 \xrightarrow{1.894} \text{CH} \xrightarrow{1.447} \text{CH} \xrightarrow{1.894} \text{CH}_2 \quad (17.56)$$

The central bond has some double-bond character, and the end bonds have some single-bond character. (In VB theory, this is explained by contributions from such

resonance structures as $\bar{\text{C}}\text{H}_2\text{—CH=CH—}\overset{+}{\text{C}}\text{H}_2$.) The sum of the bond orders is 5.235, exceeding 5. [This is because the π -electron energy of butadiene exceeds that of two isolated double bonds; see Eq. (17.58).] For benzene each carbon–carbon bond order is found to be $5/3 = 1.667$.

Naturally, we expect a relation between the bond order and the bond length R_{rs} for carbon–carbon bonds. The simplest assumption is a linear relation: $R_{rs} = a + bp_{rs}^{\text{tot}}$, where a and b are constants. The π -electron MO coefficients for ethylene and for benzene are determined completely by symmetry and are independent of the Hückel approximations. We therefore use the bond orders and bond lengths of ethylene (2; 1.335 Å) and benzene (5/3; 1.397 Å) to find a and b ; we get

$$R_{rs} = (1.707 - 0.186p_{rs}^{\text{tot}}) \text{ Å} = (1.521 - 0.186p_{rs}) \text{ Å} \quad (17.57)$$

Equation (17.57) works fairly well. Thus the predicted naphthalene 1–2, 2–3, 1–9, and 9–10 bond lengths (in Å), as compared with the experimental gas-phase values (in parentheses) are 1.386 (1.371), 1.409 (1.410), 1.418 (1.421), and 1.425 (1.420), respectively. Note that the rings are not regular hexagons. For comparison, MP2(FC)/cc-pVTZ naphthalene 1–2, 2–3, 1–9, and 9–10 bond lengths (from the CCCBDB database) are 1.377, 1.411, 1.415, and 1.430 Å, respectively.

A much-investigated question is whether the carbon–carbon bond lengths in very large conjugated linear and cyclic polyenes ($\text{C}_{2n}\text{H}_{2n+2}$ and $\text{C}_{4m+2}\text{H}_{4m+2}$ for m and n large) are equal or alternate in length. The HMO bond orders in a linear conjugated polyene become equal (except near the chain ends) in the limit of large n . The HMO bond orders in the cyclic polyene $\text{C}_{4m+2}\text{H}_{4m+2}$ are all equal. Hence it was formerly believed that in $\text{C}_{2n}\text{H}_{2n+2}$ and in $\text{C}_{4m+2}\text{H}_{4m+2}$ the bond lengths are all equal, except for those near the chain ends in $\text{C}_{2n}\text{H}_{2n+2}$. However, HMO calculations that take into account the strain energy involved in changing the σ bond lengths indicate strongly that for large n and m both linear and cyclic conjugated polyenes should have bonds that alternate in length (*Salem*, Section 8-4). Because of the crudity of the HMO method, these conclusions are not definitive.

The polymer *trans*-polyacetylene, *trans*-(CH)_x contains a very long chain of alternating single and double carbon–carbon bonds. The term *trans* refers to the configurations of the two hydrogens bonded to each pair of doubly bonded carbons (**Fig. 17.7**). Analysis of the ¹³C NMR spectrum of *trans*-(CH)_x film gave 1.36 Å for the carbon–carbon double-bond length and 1.44 Å for the single-bond length, thereby proving the existence of bond-length alternation in long-chain polyenes [C. S. Yannoni and T. C. Clarke, *Phys. Rev. Lett.*, **51**, 1191 (1983)]. A density-functional B3LYP/6-31G* geometry optimization of the linear polyene *trans*-C₃₀H₃₂ gave 1.368 and 1.426 Å for the double- and single-bond lengths in the center of the molecule [C. Choi et al., *J. Chem. Phys.*, **107**, 6712 (1997)].

B3LYP/6-31G* calculations on the cyclic polyenes $\text{C}_{4m+2}\text{H}_{4m+2}$ up to C₆₆H₆₆ predict bond-length alternation for large m [C. H. Choi and M. Kertesz, *J. Chem. Phys.*, **108**, 6681 (1998)]. The fact that we can write two equivalent VB resonance structures for $\text{C}_{4m+2}\text{H}_{4m+2}$ such that a given bond is single in one structure and double in the other is no guarantee of equal bond lengths.

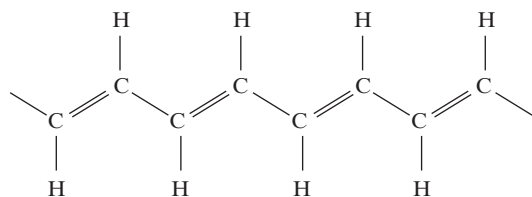


FIGURE 17.7
Trans-polyacetylene.

Bond-length alternation in the annulenes allows for greater overlap between pairs of $2p\pi$ orbitals of doubly bonded carbons due to the shorter double-bond distance, thereby stabilizing the molecule. Bond-length equality gives energy lowering due to delocalization. As noted in the discussion of Hess–Schaad resonance energy, the stabilization gained by π -electron delocalization becomes negligible as m goes to infinity in the aromatic annulenes $C_{4m+2}H_{4m+2}$. An analysis shows that even in benzene the energy lowering due to greater overlap between $2p\pi$ orbitals in the bond-alternating structure is more important than the delocalization energy gained in the equal-bond-length structure, and that it is only because the σ framework of benzene favors equal bond lengths that benzene has equal bond lengths [S. Shaik et al., *Chem. Rev.*, **101**, 1501 (2001); P. C. Hiberty and S. Shaik, *Theor. Chem. Acc.*, **114**, 169 (2005)].

The Hückel π -electron energy E_π is related to the π bond orders p_{rs} and π -electron charges q_r ; in fact (Prob. 17.16),

$$E_\pi = \alpha \sum_r q_r + 2\beta \sum_{s=r} p_{rs} \quad (17.58)$$

where the first sum is over the carbon atoms and the second sum is over the carbon–carbon bonds.

For conjugated species with partly filled degenerate MOs, there is an ambiguity in the q_r and p_{rs} values if the real forms of the MOs are used. For example, for $C_6H_6^-$, if we put the unpaired electron in $\phi_{4,\text{real}}$ of (17.42), then this electron contributes $\frac{1}{3}$ to q_1 , $\frac{1}{12}$ to q_2 , $\frac{1}{12}$ to q_3 , $\frac{1}{3}$ to q_4 , $\frac{1}{12}$ to q_5 , and $\frac{1}{12}$ to q_6 ; but if we put the odd electron in $\phi_{5,\text{real}}$, then this electron contributes 0 to q_1 and q_4 and $\frac{1}{4}$ to each of q_2 , q_3 , q_5 , and q_6 . This ambiguity can be avoided by using the complex MOs, which have the symmetry of the molecule. Putting the odd electron in either ϕ_4 or ϕ_5 of (17.41), we get a contribution of $\frac{1}{6}$ to each q . Alternatively, we can average the contributions from ϕ_4 and ϕ_5 to give a $\frac{1}{6}$ contribution at each carbon.

The HMO quantities q_r and p_{rs} in (17.53) and (17.54) are closely related to the density matrix elements P_{rs} [Eq. (14.42)] of SCF theory. We see that the density matrix elements with $r = s$ (the diagonal elements) are equal to q_r ; $P_{rr} = q_r$. Although HMO theory defines bond orders only for pairs of bonded atoms, if we formally define p_{rs} by (17.54) for non-bonded pairs of atoms also, then the definitions give $p_{rs} = \frac{1}{2}(P_{rs} + P_{sr})$, which becomes $p_{rs} = P_{rs}$ for real MOs.

Heteroatomic Conjugated Molecules

We have applied the HMO method to hydrocarbons only. For planar conjugated molecules that involve π bonding to noncarbon atoms, the α and β integrals for the heteroatoms must be modified from the carbon values. For heteroatoms X and Y, one writes

$$\alpha_X = \alpha_C + h_X \beta_{CC} \quad \text{and} \quad \beta_{XY} = k_{XY} \beta_{CC}$$

where h_X and k_{XY} are certain constants. The best values for these constants vary, depending on which molecular property is being considered, and the values used are based on a mixture of theory and guesswork. For details, see *Streitwieser*, Chapter 5.

Inclusion of Overlap

Aside from using the one-electron Hamiltonian (17.3), perhaps the most serious approximation of the simple HMO method is that of taking all overlap integrals equal to zero. Wheland proposed using a common nonzero value S for the overlap integral of carbons bonded to each other. This replaces each element β in the HMO secular equation (17.10) with $\beta - Se_i$. At the benzene carbon–carbon bond distance, S equals 0.25 for $2p\pi$ STOs with orbital exponent 1.625 (the value given by Slater's rules; Prob. 15.62). Inclusion of overlap in the HMO method is easy (see Prob. 17.24) but is rarely done. One finds that

inclusion of overlap gives only slight changes in predicted transition frequencies of alternant hydrocarbons and gives no change in the π -electron charges and bond orders when these are calculated with suitably modified definitions [B. H. Chirgwin and C. A. Coulson, *Proc. Roy. Soc.*, **A201**, 196 (1950)].

Matrix Formulation

The HMO equations (17.9) have the same form as the Roothaan equations (14.34), where H_{rs}^{eff} corresponds to F_{rs} , and e_i corresponds to ε_i . We showed the matrix form of the Roothaan equations to be $\mathbf{FC} = \mathbf{SCe}$ [Eq. (14.56)]. Hence the HMO equations are equivalent to the matrix equation $\mathbf{H}^{\text{eff}}\mathbf{C} = \mathbf{SCe} = \mathbf{Ce}$, where $\mathbf{H}^{\text{eff}}$, $\mathbf{C}$, $\mathbf{S}$, and $\mathbf{e}$ are square matrices of order n_C with elements H_{rs}^{eff} , c_{si} , $S_{rs} = \delta_{rs}$, and $e_{mi} = \delta_{mi}\varepsilon_i$. Since $\mathbf{H}^{\text{eff}}$ is real and symmetric, we have $\mathbf{C}^T\mathbf{H}^{\text{eff}}\mathbf{C} = \mathbf{e}$. Computer programs to do HMO calculations find the orthogonal matrix $\mathbf{C}$ that diagonalizes $\mathbf{H}^{\text{eff}}$. You can do online HMO calculations at www.chem.ucalgary.ca/SHMO/ and at www.stolaf.edu/depts/chemistry/courses/toolkits/247/js/huckel.

Summary

Because of the simplicity of the method, carrying out HMO calculations became a favorite pastime of organic chemists, and the results of HMO calculations have been tabulated for hundreds of compounds. The HMO method was widely used to rationalize and predict the properties and reactivities of conjugated compounds.

The development of semiempirical theories more sophisticated than the HMO theory led some workers to argue that “Hückel theory has largely outlived its usefulness” (*Murrell and Hargett*, page v). However, these more sophisticated theories have their failings (as we shall see), and the success of the Hess–Schaad use of Hückel theory to predict aromaticity indicates that HMO theory may still be useful, “especially on a qualitative level as a guide . . . in planning and interpreting experiments” (N. Trinajstić in *Segal*, Part A, Chapter 1).

17.3 The Pariser–Parr–Pople Method

Although the Hückel theory can be used to predict the longest-wavelength bands of aromatic hydrocarbons, it would be hopeless to try to use HMO theory to predict the complete electronic spectrum of an aromatic hydrocarbon. For example, Hückel theory, which neglects interelectronic repulsions, gives no separation between singlet and triplet electronic terms arising from the same configuration. Experimentally, separations of 1 or 2 eV are observed between such terms.

A semiempirical π -electron theory that takes electron repulsion into account and thereby improves on the Hückel method is the **Pariser–Parr–Pople (PPP) method**, developed in 1953. Here, the π -electron Hamiltonian (17.1) including electron repulsions is used, and the π -electron wave function is written as an antisymmetrized product of π -electron spin-orbitals. A minimal basis set of one $2p\pi$ STO on each conjugated atom is used, and the spatial π MOs ϕ_i are taken as linear combinations of these AOs: $\phi = \sum_{r=1}^b c_{ri}f_r$. The Roothaan equations are used to find SCF π MOs within the π -electron approximation.

The π -electron Hamiltonian $\hat{H}_\pi$ in (17.1) has the same form as the all-electron operator $\hat{H}_{\text{el}} = -\frac{1}{2}\sum_i \nabla_i^2 - \sum_i \sum_\alpha Z_\alpha/r_{i\alpha} + \sum_i \sum_{j>i} 1/r_{ij}$ for a molecule except that $\hat{H}_\pi^{\text{core}}(i)$ replaces $\hat{H}^{\text{core}}(i) \equiv -\frac{1}{2}\nabla_i^2 - \sum_\alpha Z_\alpha/r_{i\alpha}$ [Eq. (14.27)], and the sums go over only the n_π π electrons rather than over all the electrons. Hence, similar to (14.25), the SCF π MOs satisfy $\hat{F}_\pi\phi_i = \varepsilon_i\phi_i$, where $\hat{F}_\pi$ is given by (14.26) with $\hat{H}^{\text{core}}(1)$ replaced by $\hat{H}_\pi^{\text{core}}$ and $n/2$ replaced by $n_\pi/2$. The SCF π -electron energy is given by (14.22) with H_{ii}^{core} replaced by $H_{\pi,ii}^{\text{core}}$ and $n/2$ replaced by $n_\pi/2$.

The Roothaan equations (14.34) and (14.41) become

$$\sum_s c_{si} (F_{\pi,rs} - \varepsilon_i S_{rs}) = 0, \quad r = 1, \dots, b \quad (17.59)$$

$$F_{\pi,rs} = H_{\pi,rs}^{\text{core}} + \sum_{t=1}^b \sum_{u=1}^b P_{tu} [(rs|tu) - \frac{1}{2}(ru|ts)] \quad (17.60)$$

In addition to assuming σ - π separability, the PPP method makes further approximations. As in Hückel theory, overlap is neglected:

$$S_{rs} \equiv \langle f_r(1) | f_s(1) \rangle = \delta_{rs} \quad (17.61)$$

where δ_{rs} is the Kronecker delta. Consistent with the neglect of overlap integrals, when evaluating electron-repulsion integrals the PPP method makes the approximation of **zero differential overlap** (ZDO):

$$[f_r(1)]^* f_s(1) dv_1 = 0, \quad \text{for } r \neq s \quad (17.62)$$

From (17.62) and $(rs|tu) = \langle f_r(1)f_t(2) | 1/r_{12} | f_s(1)f_u(2) \rangle$ [Eq. (14.39)], it follows that the electron-repulsion integrals are given by

$$(rs|tu) = \delta_{rs}\delta_{tu}(rr|tt) = \delta_{rs}\delta_{tu}\gamma_{rt} \quad (17.63)$$

where $\gamma_{rt} \equiv (rr|tt)$. Thus the method ignores many (but not all) of the electron-repulsion integrals, thereby greatly simplifying the calculation. In particular, all three- and four-center electron-repulsion integrals are ignored. The ZDO approximation is not used in the $H_{\pi,rs}^{\text{core}}$ integrals.

The ZDO approximation is at first sight quite drastic. However, a partial theoretical justification for it can be given by reinterpreting the AOs used to express the MOs as orthogonalized AOs (rather than ordinary AOs). Each orbital in a set of orthogonalized AOs is a linear combination of ordinary AOs, the coefficients being chosen so that the members of the set are mutually orthogonal. There are many ways to choose the linear combinations to produce an orthogonal set. One approach is to make the orthogonalized AOs (OAOs) resemble the ordinary AOs as much as possible by minimizing the sum of the squares of the deviations of the OAOs from the ordinary AOs. One minimizes the sum $\sum_i \int |\chi_{i,\text{OAO}} - \chi_i|^2 dv$, where the sum goes over the set of AOs and where χ_i and $\chi_{i,\text{OAO}}$ are the ordinary and the orthogonalized AOs. This produces what are called *symmetrically orthogonalized* (or Löwdin) AOs (see Prob. 8.58; *Pilar*, Section 14-8). One finds that in each symmetrically orthogonalized AO the coefficient of one ordinary AO is substantially greater than the coefficients of the other ordinary AOs, so the symmetrically orthogonalized AOs are not drastically different from the ordinary AOs. Moreover, with symmetrically orthogonalized AOs the integrals not neglected in the ZDO approximation undergo only small changes in value as compared with their values with ordinary AOs. These changes in value can be partly allowed for by the fact that many integrals are taken as empirical parameters. Also, with symmetrically orthogonalized AOs, the electron-repulsion integrals neglected in the ZDO approximation are generally found to be quite small (and all overlap integrals are, of course, zero). For details, see the references cited on page 27 of *Murrell and Harget*.

With the approximation (17.63) for $(rs|tu)$, the matrix elements $F_{\pi,rs}$ in (17.60) become (Prob. 17.25)

$$\begin{aligned} F_{\pi,rr} &= H_{\pi,rr}^{\text{core}} + \sum_t P_{tt}\gamma_{rt} - \frac{1}{2}P_{rr}\gamma_{rr} \\ F_{\pi,rs} &= H_{\pi,rs}^{\text{core}} - \frac{1}{2}P_{sr}\gamma_{rs}, \quad r \neq s \end{aligned} \quad (17.64)$$

The PPP method does not attempt to explicitly specify $\hat{H}_{\pi}^{\text{core}}$ or to calculate the $H_{\pi,rs}^{\text{core}}$ integrals theoretically. Rather, the integrals $H_{\pi,rs}^{\text{core}}$ and γ_{rs} are calculated from approximate semiempirical formulas, some of which contain empirical parameters. For example, when the AOs f_r and f_s are on atoms R and S that are bonded to each other, $H_{\pi,rs}^{\text{core}}$ may be taken as $k\langle f_r|f_s\rangle$, where the value of the empirical parameter k is chosen so that the predictions of the theory give good agreement with experiment; the overlap integral $\langle f_r|f_s\rangle$ is calculated from the STOs f_r and f_s , and not taken as zero as in (17.62). When the two different atoms R and S are not bonded to each other, $H_{\pi,rs}^{\text{core}}$ is taken as zero. (Several versions of the PPP theory exist, each of which uses a different set of semiempirical formulas to evaluate the integrals.) The two-center electron-repulsion integrals γ_{rs} are evaluated from a semiempirical formula that contains the one-center integrals γ_{rr} and γ_{ss} and the distance between atoms R and S. The one-center integral γ_{rr} is evaluated as the difference between the atomic valence-state ionization energy and electron affinity of atom R, where these two quantities are evaluated using spectroscopically determined atomic energy-level data. The one-center integral $H_{\pi,rr}^{\text{core}}$ is calculated from a semiempirical formula that involves the electron-repulsion integrals γ_{rs} for all the conjugated atoms $S \neq R$ and involves the orbital energy for a $2p\pi$ AO on atom R, this orbital energy being evaluated using atomic spectral data.

To do a PPP calculation, one starts with the HMO coefficients as an initial guess for the c_{si} 's, calculates the initial density matrix elements P_{rs} , calculates the initial $F_{\pi,rs}$ matrix elements, solves the equations (17.59) for π -electron orbital energies ε_i and an improved set of coefficients c_{si} , calculates improved P_{rs} values, and so on until convergence is reached. To improve the results, CI of the π electrons may be included.

The PPP method gives a good account of the electronic spectra of many, but not all, aromatic hydrocarbons. For more on the PPP method, see *Parr*, Chapter III; *Murrell and Harget*, Chapter 2; *Offenhartz*, Chapter 11.

The PPP method is not used nowadays and has been superseded by more general semiempirical methods (Section 17.4). However, the PPP method is of historical importance, since many of the PPP approximations used to evaluate integrals are used in current semiempirical theories.

17.4 General Semiempirical MO and DFT Methods

The HMO and PPP methods apply only to planar conjugated molecules and treat only the π electrons. The semiempirical MO methods discussed in this section apply to all molecules and treat all the valence electrons.

Semiempirical MO theories fall into two categories: those using a Hamiltonian that is the sum of one-electron terms, and those using a Hamiltonian that includes two-electron repulsion terms, as well as one-electron terms. The Hückel method is a one-electron theory, whereas the Pariser–Parr–Pople method is a two-electron theory.

The Extended Hückel Method

The most important one-electron semiempirical MO method for nonplanar molecules is the **extended Hückel theory**. An early version was used by Wolfsberg and Helmholz in treating inorganic complex ions. The method was further developed and widely applied by Hoffmann [R. Hoffmann, *J. Chem. Phys.*, **39**, 1397 (1963); **40**, 2745, 2474, 2480 (1964); *Tetrahedron*, **22**, 521, 539 (1966); M. Wolfsberg and L. Helmholz, *J. Chem. Phys.*, **20**, 837 (1952)].

The extended Hückel (EH) method begins with the approximation of treating the valence electrons separately from the rest (Section 13.17). The valence-electron Hamiltonian is taken as the sum of one-electron Hamiltonians:

$$\hat{H}_{\text{val}} = \sum_i \hat{H}_{\text{eff}}(i) \quad (17.65)$$

where $\hat{H}_{\text{eff}}(i)$ is not specified explicitly. The MOs are approximated as linear combinations of the valence AOs f_r of the atoms:

$$\phi_i = \sum_{r=1}^b c_{ri} f_r \quad (17.66)$$

In the simple Hückel theory of planar hydrocarbons, each π MO contains contributions from one $2p\pi$ AO on each carbon atom. In the extended Hückel treatment of nonplanar hydrocarbons, each valence MO contains contributions from four AOs on each carbon atom (one $2s$ and three $2p$'s) and one $1s$ AO on each hydrogen atom. The AOs used are usually Slater-type orbitals with fixed orbital exponents determined from Slater's rules (Prob. 15.62). For the simplified Hamiltonian (17.65), the problem separates into several one-electron problems:

$$\begin{aligned} \hat{H}_{\text{eff}}(i)\phi_i &= e_i\phi_i \\ E_{\text{val}} &= \sum_i e_i \end{aligned} \quad (17.67)$$

Application of the variation theorem to the linear trial function (17.66) gives as the secular equation and the equations for the MO coefficients

$$\det(H_{rs}^{\text{eff}} - e_i S_{rs}) = 0 \quad (17.68)$$

$$\sum_s [(H_{rs}^{\text{eff}} - e_i S_{rs})c_{si}] = 0, \quad r = 1, 2, \dots, b \quad (17.69)$$

All this is similar to simple Hückel theory. However, the extended Hückel theory does not neglect overlap. Rather, *all* overlap integrals are explicitly evaluated using the forms chosen for the AOs and the internuclear distances at which the calculation is being done. Formulas for overlap integrals of STOs are readily available. [Footnotes 12 to 18 of D. M. Bishop et al., *J. Chem. Phys.*, **45**, 1880 (1966) give references to available tabulations of overlap integrals.] Since overlap is included, off-diagonal e_i 's are present in the secular determinant.

Since $\hat{H}_{\text{eff}}(i)$ is not specified, there is the problem of what to use for the integrals H_{rs}^{eff} . For $r = s$, the one-electron integral $H_{rr}^{\text{eff}} \equiv \langle f_r | \hat{H}_{\text{eff}} | f_r \rangle$ looks like an average energy for an electron in the AO f_r centered on atom R in the molecule. Hence, the EH method takes H_{rr}^{eff} as equal to the orbital energy of the AO f_r for atom R in its valence state; the *valence state* is the hypothetical state of the atom in the molecule. The valence-state orbital energy can be found from atomic spectral data. By Koopmans' theorem (Section 15.5), the valence-state orbital energy is taken as equal to minus the valence-state ionization potential (VSIP) of f_r . For hydrogen and for carbon atoms in a molecule with only single bonds to carbon (sp^3 -hybridized carbon), the VSIP parametrization gives

$$\begin{aligned} \langle C2s | \hat{H}_{\text{eff}} | C2s \rangle &= -20.8 \text{ eV}, & \langle C2p | \hat{H}_{\text{eff}} | C2p \rangle &= -11.3 \text{ eV} \\ \langle H1s | \hat{H}_{\text{eff}} | H1s \rangle &= -13.6 \text{ eV} \end{aligned} \quad (17.70)$$

The carbon VSIPs for sp^2 - and sp -hybridized carbon differ from those in (17.70), but the difference is often ignored and an average set of VSIPs is used for all carbons. VSIPs are tabulated in J. Hinze and H. H. Jaffé, *J. Am. Chem. Soc.*, **84**, 540 (1962); G. Pilcher and H. A. Skinner, *J. Inorg. Nuc. Chem.*, **24**, 937 (1962); L. C. Cusachs et al., *J. Chem. Phys.*, **44**, 835 (1966).

For the off-diagonal matrix elements H_{rs}^{eff} , $r \neq s$, Wolfsberg, Helmholz, and Hoffmann took

$$H_{rs}^{\text{eff}} = \frac{1}{2} K (H_{rr}^{\text{eff}} + H_{ss}^{\text{eff}}) S_{rs} \quad (17.71)$$

where K is a numerical constant and the remaining quantities are evaluated as above. Since H_{rr}^{eff} and H_{ss}^{eff} are usually negative, (17.71) gives H_{rs}^{eff} as negative. The most commonly used value of K is 1.75 (although values between 1 and 3 have been suggested). In contrast to the simple Hückel theory, H_{rs}^{eff} is nonzero for all pairs of orbitals unless S_{rs} vanishes for symmetry reasons.

Once the H_{rs}^{eff} and S_{rs} integrals have been evaluated, the secular equation is solved for the orbital energies and the MO coefficients are found. (Computer programs use matrix diagonalization to do this.) Since the H_{rs}^{eff} integrals do not depend on the MO coefficients, no iteration is needed. The EH method is often used to provide an initial guess for the valence MO coefficients in ab initio SCF MO calculations.

The total valence-electron energy is given by (17.67) as the sum of orbital energies. To predict molecular geometry using the EH method, one does a series of calculations over a range of bond distances and angles and looks for the nuclear configuration that minimizes (17.67). Note that (17.67) omits both electron–electron repulsions and nuclear–nuclear repulsions. It might be thought that such a theory would be useless for predicting molecular geometries, but this is not so.

The EH method gives rather accurate bond angles for molecules whose bonds are not highly polar, but fails in bond-angle predictions for molecules with very polar bonds (for example, H_2O , which is predicted to be linear). The EH method is not reliable for predicting bond lengths, dipole moments, molecular conformations, and barriers to internal rotation.

Because the EH method gives poor predictions of molecular properties, Jug concluded that this method “is obsolete” [K. Jug, *Theor. Chim. Acta*, **54**, 263 (1980)]. However, this judgment is too harsh, since Hoffmann and others have used the EH method to provide valuable qualitative insights into chemical bonding [for example, P. J. Hay, J. C. Thibault, and R. Hoffmann, *J. Am. Chem. Soc.*, **97**, 4884 (1975)]. Gimarc noted that “the real value of the extended Hückel method is not in its quantitative results, which have never been impressive, but rather in the qualitative nature of the results and in the interpretations those results can provide” (B. M. Gimarc, *Molecular Structure and Bonding*, Academic Press, 1979, page 216).

The CNDO, INDO, and NDDO Methods

Several semiempirical two-electron MO generalizations of the PPP method were developed that apply to both planar and nonplanar molecules. The **complete neglect of differential overlap (CNDO) method** was proposed by Pople, Santry, and Segal in 1965. The **intermediate neglect of differential overlap (INDO) method** was proposed by Pople, Beveridge, and Dobosh in 1967. Both methods treat only the valence electrons explicitly. The valence-electron Hamiltonian has a similar form as (17.1):

$$\hat{H}_{\text{val}} = \sum_{i=1}^{n_{\text{val}}} \left[-\frac{1}{2} \nabla_i^2 + V(i) \right] + \sum_{i=1}^{n_{\text{val}}} \sum_{j>i} \frac{1}{r_{ij}} \equiv \sum_{i=1}^{n_{\text{val}}} \hat{H}_{\text{val}}^{\text{core}}(i) + \sum_{i=1}^{n_{\text{val}}} \sum_{j>i} \frac{1}{r_{ij}} \quad (17.72)$$

$$\hat{H}_{\text{val}}^{\text{core}}(i) \equiv -\frac{1}{2} \nabla_i^2 + V(i)$$

In (17.72), n_{val} is the number of valence electrons in the molecule, $V(i)$ is the potential energy of valence electron i in the field of the nuclei and the inner-shell (core) electrons, and $\hat{H}_{\text{val}}^{\text{core}}(i)$ is the one-electron part of $\hat{H}_{\text{val}}$. The CNDO and INDO methods are SCF MO methods that iteratively solve the Roothaan equations using approximations for the integrals in the Fock matrix elements.

The CNDO method uses a minimal basis set of valence Slater AOs f_r with orbital exponents fixed at values given by Slater’s rules (Prob. 15.62), except that 1.2 is used for H1s. The valence MOs ϕ_i are written as $\phi_i = \sum_{r=1}^b c_{ri} f_r$. The molecular electronic energy

is given by (14.22) with H_{ii}^{core} replaced by $H_{\text{val},ii}^{\text{core}}$, $n/2$ replaced by $n_{\text{val}}/2$, and V_{NN} replaced by the **core–core repulsion** energy

$$V_{cc} = \sum_{\alpha} \sum_{\beta > \alpha} \frac{C_{\alpha} C_{\beta}}{R_{\alpha\beta}}$$

where the **core charge** C_{α} on atom α equals the atomic number of atom α minus the number of core (inner-shell) electrons on α . The Roothaan equations are given by (17.59) with $F_{\pi,rs}$ replaced by $F_{\text{val},rs}$. The Fock matrix elements $F_{\text{val},rs}$ [Eq. (14.41)] are given by (17.60) with $H_{\pi,rs}^{\text{core}}$ replaced by $H_{\text{val},rs}^{\text{core}} \equiv \langle f_r | \hat{H}_{\text{val}}^{\text{core}} | f_s \rangle$:

$$F_{\text{val},rs} = H_{\text{val},rs}^{\text{core}} + \sum_{t=1}^b \sum_{u=1}^b P_{tu} [(rs|tu) - \frac{1}{2}(ru|ts)] \quad (17.73)$$

Henceforth in this section, the subscript val will be omitted from $F_{\text{val},rs}$, $H_{\text{val},rs}^{\text{core}}$, and $\hat{H}_{\text{val}}^{\text{core}}(i)$.

The CNDO method uses the ZDO approximation (17.62) for all pairs of AOs in overlap and electron-repulsion integrals. Thus, $S_{rs} = \delta_{rs}$ and $(rs|tu) = \delta_{rs}\delta_{tu}(rr|tt) \equiv \delta_{rs}\delta_{tu}\gamma_{rt}$ [Eq. (17.63)], where $(rs|tu)$ is given by (14.39). In the PPP method for conjugated hydrocarbons, there is only one basis AO per atom, the $2p\pi$ AO. In the CNDO method, there are several basis valence AOs on each atom (except hydrogens), and the ZDO approximation neglects electron-repulsion integrals containing the product $f_r(1)f_s(1)$ where f_r and f_s are different AOs centered on the same atom.

When the Roothaan equations (14.34) [or (14.56)] are solved exactly, the canonical MOs and the calculated values of molecular properties do not change if one changes the orientation of the coordinate axes. The calculated values are said to be *rotationally invariant*. Likewise, the results do not change if each basis AO on a particular atom is replaced by a linear combination of the basis AOs on that atom, and the results are *hybridizationally invariant*. When approximations are made in solving the Hartree–Fock–Roothaan equations, rotational and hybridizational invariance may not hold.

To maintain rotational and hybridizational invariance when the ZDO approximation is used, the CNDO method makes the additional approximation that the electron-repulsion integral $\gamma_{rt} \equiv (rr|tt)$ depends only on which atoms the AOs f_r and f_t are centered on and does not depend on the nature of the orbitals f_r and f_t . Let the notation f_{r_A}, f_{t_B} indicate that the valence AOs f_r and f_t are centered on atoms A and B, respectively, where A and B might be the same atom or different atoms. The CNDO method takes $(r_A r_A | t_B t_B) \equiv \gamma_{r_A t_B} = \gamma_{AB}$ for all valence AOs f_r on A and all valence AOs f_t on B. In the CNDO method, all one-center valence-electron-repulsion integrals on atom A have the value γ_{AA} ; all two-center valence electron-repulsion integrals involving atoms A and B have the value γ_{AB} ; all three- and four-center electron-repulsion integrals are neglected (as a consequence of the ZDO approximation). The integrals γ_{AA} and γ_{AB} are evaluated using valence s STOs on A and B, and so they depend on the orbital exponents, the principal quantum numbers of the valence electrons, and on the distance between atoms A and B.

CNDO takes the integral H_{rs}^{core} for $r \neq s$ as proportional to the overlap integral $S_{rs} \equiv \langle f_r | f_s \rangle$ (recall a similar assumption in the PPP theory):

$$H_{r_A s_B}^{\text{core}} = \beta_{AB}^0 S_{r_A s_B} \quad \text{for } r \neq s$$

where $S_{r_A s_B}$ is evaluated exactly [and not taken as δ_{rs} , even though it is taken as δ_{rs} in the Roothaan equations (17.59)] and the parameter β_{AB}^0 is taken as $\beta_{AB}^0 = \frac{1}{2}(\beta_A^0 + \beta_B^0)$, where the parameters β_A^0 and β_B^0 are chosen so that the coefficients in CNDO-calculated MOs of diatomic molecules resemble the coefficients in minimal-basis ab initio MOs. When A and

B are the same atom, $S_{r_A s_A}$ is zero for $r \neq s$ by orthogonality of AOs on the same atom, and CNDO takes $H_{r_A s_A}^{\text{core}}$ as zero for $r \neq s$.

Now consider the integrals $H_{r_A r_A}^{\text{core}}$. We have $\hat{H}^{\text{core}}(1) = -\frac{1}{2}\nabla_1^2 + V(1)$, where $V(1)$ is the potential energy of valence electron 1 in the field of the core electrons and the nuclei. Breaking $V(1)$ into contributions from the individual atomic cores, we have

$$\hat{H}^{\text{core}}(1) = -\frac{1}{2}\nabla_1^2 + V_A(1) + \sum_{B \neq A} V_B(1) \quad (17.74)$$

where the basis AO f_{r_A} is centered on A. Then

$$H_{r_A r_A}^{\text{core}} = \langle f_{r_A}(1) | -\frac{1}{2}\nabla_1^2 + V_A(1) | f_{r_A}(1) \rangle + \sum_{B \neq A} \langle f_{r_A}(1) | V_B(1) | f_{r_A}(1) \rangle \quad (17.75)$$

Two versions of CNDO exist, called CNDO/1 and CNDO/2. The integral

$$U_{rr} \equiv \langle f_{r_A}(1) | -\frac{1}{2}\nabla_1^2 + V_A(1) | f_{r_A}(1) \rangle$$

looks like an average energy for an electron in the AO f_{r_A} in the molecule. Hence, CNDO/1 takes U_{rr} as the negative of the valence-state ionization energy from the AO f_{r_A} , where this ionization energy is found from atomic energy levels deduced from atomic spectral data. (This is an oversimplification; for details and for the CNDO/2 evaluation of U_{rr} , see *Murrell and Harget*, pp. 38–39.) To maintain rotational and hybridizational invariance, the integrals $\langle f_{r_A}(1) | V_B | f_{r_A}(1) \rangle \equiv V_{AB}$ in (17.75) are taken as equal for all valence AOs f_{r_A} on atom A and are calculated in CNDO/1 by using a valence s orbital for electron 1 and taking atomic core B as a point charge: $V_{AB} = -\langle s_A(1) | C_B/r_{1B} | s_A(1) \rangle$, where C_B is the core charge of atom B. With this choice for the electron–core attraction integral V_{AB} in CNDO/1, two neutral atoms or molecules separated by several angstroms experienced a substantial attraction to each other. To eliminate this spurious attraction, CNDO/2 uses the expression $V_{AB} = -C_B\gamma_{AB}$, where γ_{AB} is the electron–repulsion integral discussed earlier in this subsection.

With these approximations, the Fock matrix elements are evaluated and the Roothaan equations are solved iteratively to find the CNDO orbitals and orbital energies.

The INDO method is an improvement on CNDO. In INDO, differential overlap between AOs on the same atom is not neglected in one-center electron–repulsion integrals $(rs|tu)$ where f_r, f_s, f_t , and f_u are all centered on the same atom, but is still neglected in two-center electron–repulsion integrals. Thus fewer two-electron integrals are neglected than in CNDO. Otherwise, the two methods are essentially the same. The INDO method gives an improvement on CNDO results, especially where electron spin distribution is important (for example, in calculating electron–spin–resonance spectra).

As to results, the CNDO and INDO methods give fairly good bond lengths and angles, somewhat erratic dipole moments, and poor dissociation energies. (For details on CNDO and INDO, see *Pople and Beveridge*; *Leach*, Section 2.9; G. Klopman and R. C. Evans in *Segal*, Part A, page 29; *Murrell and Harget*, Chapter 3.)

Versions of CNDO and INDO parametrized to predict electronic spectra are called CNDO/S and INDO/S. These methods include some configuration interaction. Although the ground state of a closed-shell molecule is generally well represented by a single-determinant wave function, one typically requires CI for accurate representation of excited states. For details of these methods, see R. L. Ellis and H. H. Jaffé in *Segal*, Part B, page 49; J. Michl in *Segal*, Part B, page 99.

The CNDO and INDO methods are not used nowadays, since they have been made obsolete by the improved semiempirical methods discussed in the next subsection. The one exception to this statement is INDO/S, which is still widely used to calculate electronic

spectra, since it gives good results for vertical excitation energies (Section 16.2) of large molecules, including transition-metal compounds [see M. C. Zerner in K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry*, vol. 2, VCH, 1991, pp. 333–335, 348–353]; INDO/S is often called ZINDO, after the ZINDO program of Zerner.

The **neglect of diatomic differential overlap** (NDDO) method (suggested by Pople, Santry, and Segal in 1965) is an improvement on INDO in which differential overlap is neglected only between AOs centered on different atoms: $f_r^*(1)f_s(1) dv_1 = 0$ only when AOs r and s are on different atoms. The degree of neglect of differential overlap in NDDO is more justifiable than in CNDO or INDO. The NDDO method satisfies the rotational and hybridizational invariance conditions without the need to use a common value for electron-repulsion integrals involving different valence AOs on a given atom.

A few initial attempts at parametrizing the NDDO method gave rather disappointing results (see G. Klopman and R. C. Evans in *Segal*, Part A, Chapter 2), and the method was little used until 1977, when Dewar and Thiel modified it to create the MNDO method, discussed later in this section.

The MNDO, AM1, PM3, PM6, PM6-D3H4, PM7, and RM1 Methods

Pople's aim in the CNDO and INDO methods was to reproduce as well as possible the results of minimal-basis-set ab initio SCF MO calculations with theories requiring much less computer time than ab initio calculations. Since CNDO and INDO use approximations, we can expect their results to be similar to but less accurate than minimal-basis ab initio SCF MO results. Thus these methods do pretty well on molecular geometry but fail for binding energies. Dewar and co-workers devised several semiempirical SCF MO theories that closely resemble the INDO and NDDO methods. However, Dewar's aim was not to reproduce ab initio SCF wave functions and properties but to have a theory that would give molecular binding energies with chemical accuracy (within 1 kcal/mol) and that could be used for large molecules without a prohibitive amount of calculation. It might seem unlikely that one could devise an SCF MO theory that involves approximations to the ab initio Hartree–Fock method but that succeeds for binding energies, where the Hartree–Fock theory fails. However, by proper choice of the parameters in the semiempirical SCF theory, one can actually get better results than ab initio SCF calculations, because the choice of suitable parameters can compensate for the partial neglect of electron correlation in ab initio SCF theory.

The semiempirical theories of this subsection, which follow Dewar's approach to parametrization, will be called *Dewar-type theories*. These theories treat only the valence electrons, and most of these theories use a minimal-basis set of valence Slater-type s and p AOs (with orbital exponents given values determined by parametrization) to expand the valence-electron MOs. (Extensions to include atoms with valence d orbitals are discussed later in this subsection.) The Fock–Roothaan equations (with the overlap integrals S_{rs} taken as δ_{rs}) are solved to find semiempirical SCF MOs. Some degree of neglect of differential overlap is used to eliminate many of the electron-repulsion integrals. In ab initio methods, the integrals occurring in the Fock matrix elements F_{rs} are evaluated accurately, but this is not the approach used in Dewar-type theories. Dewar-type theories take the one-center electron-repulsion integrals (ERIs) as parameters whose values are chosen to fit experimental atomic energy-level data and calculate the two-center ERIs from the values of the one-center ERIs and the internuclear distances using an approximate formula that may involve parameters. The remaining integrals are evaluated from approximate parameter-containing formulas that are designed not to give values that accurately reproduce ab initio values but to be consistent with the approximations used in the theory.

The Dewar-type theories are parametrized so as to yield good values of the 25°C gas-phase standard enthalpy of formation $\Delta H_{f,298}^\circ$. These theories calculate $\Delta H_{f,298}^\circ$ as follows.

The molecular valence electronic energy U_{val} including nuclear repulsion is taken as the sum of the purely electronic energy $E_{\text{el, val}}$ of the valence electrons (which is found from the results of the semiempirical SCF calculation) and the **core–core repulsion energy** V_{cc} [compare Eq. (13.8)]:

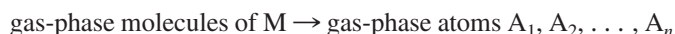
$$U_{\text{val}} = E_{\text{el, val}} + V_{cc}$$

The Dewar-type theories treat the molecule as a collection of valence electrons and atomic cores, where each core consists of an atomic nucleus and the inner-shell (core) electrons. For example, the core of a carbon atom consists of the nucleus and the two 1s electrons. The simplest approach would be to take $V_{cc} = \sum_{B>A} \sum_A V_{cc, AB} = \sum_{B>A} \sum_A C_A C_B / R_{AB}$, where C_A and C_B are the core charges of cores A and B. For example, for a carbon atom, $C_A = 6 - 2 = 4$, the number of valence electrons. Although this form of V_{cc} is used in CNDO and INDO, it is more consistent with the approximations used to evaluate the electron–core interaction integrals in Dewar-type theories to take

$$V_{cc} = \sum_{B>A} \sum_A [C_A C_B (s_A s_A | s_B s_B) + f_{AB}] \quad (17.76)$$

where the electron-repulsion integral $(s_A s_A | s_B s_B)$ involves the valence s orbitals of atoms A and B (and is approximately proportional to $1/R_{AB}$), and f_{AB} is a small term whose form differs in the various theories. f_{AB} is an empirical function of R_{AB} that fine-tunes interatomic attractions and repulsions in the molecule, so as to improve agreement with experiment.

The equilibrium geometry is then found by minimizing U_{val} (Section 15.10) to give the equilibrium valence electronic energy including core repulsion: $U_{\text{val, e}} = E_{\text{el, val, e}} + V_{cc, e}$. Let molecule M be composed of the atoms $A_1, A_2, \dots, A_n$. The molecular dissociation (or atomization) energy $D_{e, M}$ of M (Section 15.13) is calculated as $D_{e, M} = \sum_i E_{\text{val, } A_i} - U_{\text{val, e, M}}$, where $E_{\text{val, } A_i}$ is the valence electronic energy of atom A_i calculated using the same Dewar-type method as used to calculate $U_{\text{val, e, M}}$, and the sum goes over all atoms of the molecule. For the dissociation reaction



we can write that at 25°C, $\Delta H_{f, 298}^\circ = \sum_i \Delta H_{f, 298, A_i(g)}^\circ - \Delta H_{f, 298, M(g)}^\circ$, where $\Delta H_{f, 298, M(g)}^\circ$ and $\Delta H_{f, 298, A_i(g)}^\circ$ are the *gas-phase* 25°C standard enthalpies of formation of the molecular species M and of atom A_i . At 0 K, the enthalpy and energy changes for the dissociation reaction are equal, and if we ignore zero-point energy, we can write $\Delta H_0^\circ = N_A D_e$, where the Avogadro constant N_A converts from a per molecule to a per mole basis. If we ignore the difference between 0 K and 298 K enthalpy changes and take $\Delta H_0^\circ = \Delta H_{f, 298}^\circ$, we have $N_A D_e = \sum_i \Delta H_{f, 298, A_i(g)}^\circ - \Delta H_{f, 298, M(g)}^\circ$. Using the preceding expression for D_e , we get the equation used to calculate a gas-phase $\Delta H_{f, 298}^\circ$ value in a Dewar-type theory:

$$\Delta H_{f, 298, M(g)}^\circ = N_A U_{\text{val, e, M}} - N_A \sum_i E_{\text{val, } A_i} + \sum_i \Delta H_{f, 298, A_i(g)}^\circ$$

where $U_{\text{val, e}} = E_{\text{el, val, e}} + V_{cc, e}$ and the $\Delta H_{f, 298, A_i(g)}^\circ$ values are taken from thermodynamics tables. This procedure ignores the zero-point vibrational energy (which should be subtracted from D_e to give D_0) and ignores the enthalpy changes from 0 K to 298 K (Section 15.13). The justification for ignoring these nonnegligible quantities is that the parameters of the theory are chosen to fit data that include experimental $\Delta H_{f, 298}^\circ$ values of many compounds, so the parameter values include built-in corrections for the neglected quantities.

The parameter values in a Dewar-type theory are chosen as follows. One decides on a set of elements (for example, C, H, O, N, P, Si, S, F, Cl, Br, I) for which the theory is to

be parametrized, and one chooses a set of molecules (containing only these elements) for which $\Delta H_{f,298}^\circ$ and the molecular geometry and dipole moment are known from experiment. One systematically varies the parameters of the theory so as to minimize the errors in the calculated heats of formation, geometries, and dipole moments.

This process resembles the process of optimizing a molecular geometry (Section 15.10). In geometry optimization, one varies the bond distances, angles, and dihedral angles so as to minimize the molecular electronic energy, including nuclear repulsion. In a Dewar-type-theory parametrization, one varies the parameters' values so as to minimize the weighted sums of the squares of the errors in calculated molecular properties Y_i —that is, to minimize $\sum_i W_i (Y_{i,\text{calc}} - Y_{i,\text{exper}})^2$, where the W_i 's are weighting factors that determine the relative importance of the properties Y_i in the process. As in geometry optimization, it is hard to be sure the global minimum has been found (Section 15.11).

The first useful Dewar-type theory was the MINDO/3 (third version of the modified INDO) method, published in 1975. For a sample of compounds containing no elements other than C, H, O, and N, the average absolute errors in MINDO/3 calculated properties are 11 kcal/mol in heats of formation, 0.022 Å in bond lengths, 5.6° in bond angles, 0.49 D in dipole moments, and 0.7 eV in ionization energies. The errors in $\Delta H_{f,298}^\circ$ are larger than Dewar was aiming for.

MINDO/3 is based on the INDO approximation. The remaining theories of this subsection are based on the far more justifiable NDDO approximation (see the preceding subsection). These NDDO-based Dewar-type theories give significantly better results than MINDO/3, and MINDO/3 is no longer used.

Because MINDO/3 did not meet Dewar's aims, Dewar and Thiel developed the **MNDO** (modified neglect of diatomic overlap) method. The MNDO method has been parametrized for nearly all the main-group elements and for Zn, Cd, and Hg. MNDO gives substantially improved results as compared with MINDO/3. For the same sample of C, H, O, N compounds used above for MINDO/3 errors, average absolute MNDO errors are 6.3 kcal/mol in heats of formation, 0.014 Å in bond lengths, 2.8° in bond angles, 0.30 D in dipole moments, and 0.5 eV in ionization energies [M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **99**, 4907 (1977)].

The MNDO valence-electron Hamiltonian $\hat{H}_{\text{val}}$ is given by (17.72), and the Fock matrix elements are given by (17.73). As in (17.74), the core Hamiltonian operator for valence electron 1 is written as $\hat{H}^{\text{core}}(1) = -\frac{1}{2}\nabla_1^2 + \sum_B V_B(1)$, where $V_B(1)$ is the part of the potential energy of electron 1 due to its interactions with the core (nucleus plus inner-shell electrons) of atom B. To evaluate the Fock matrix element $F_{\mu\nu}$ in (17.73), we need the core matrix elements $H_{\mu\nu}^{\text{core}}$, the electron-repulsion integrals $(\mu\nu|\lambda\sigma)$, and an initial guess for the density matrix elements $P_{\lambda\sigma}$, so as to start the SCF iterative process. Here $\mu, \nu, \lambda, \sigma$ are used instead of r, s, t, u as subscripts to avoid confusion with the use of s to denote an s orbital.

The integrals occurring in the MNDO Fock matrix elements $F_{\mu\nu}$ are evaluated as follows.

The core matrix element $H_{\mu_A\nu_B}^{\text{core}} = \langle \mu_A(1) | \hat{H}^{\text{core}}(1) | \nu_B(1) \rangle$ (often called a core resonance integral) involving AOs centered on different atoms A and B is taken (as in CNDO) as proportional to the overlap integral (evaluated exactly) between these AOs:

$$H_{\mu_A\nu_B}^{\text{core}} = \frac{1}{2}(\beta_{\mu_A} + \beta_{\nu_B})S_{\mu_A\nu_B}, \quad A \neq B \quad (17.77)$$

where the β 's are parameters for each type of valence AO. For example, carbon has valence AOs of types $2s$ and $2p$, and MNDO has parameters β_{C2s} and β_{C2p} .

The core matrix element $H_{\mu_A\nu_A}^{\text{core}} = \langle \mu_A(1) | \hat{H}^{\text{core}}(1) | \nu_A(1) \rangle$ involving different AOs centered on the same atom is evaluated using $\hat{H}^{\text{core}}(1) = -\frac{1}{2}\nabla_1^2 + V_A(1) + \sum_{B \neq A} V_B(1)$ [Eq. (17.74)] to write

$$H_{\mu_A\nu_A}^{\text{core}} = \langle \mu_A | -\frac{1}{2}\nabla^2 + V_A | \nu_A \rangle + \sum_{B \neq A} \langle \mu_A | V_B | \nu_A \rangle$$

The integral $\langle \mu_A | -\frac{1}{2}\nabla^2 + V_A | \nu_A \rangle$ can be shown to be zero using a theorem of group theory (*Offenhardt*, p. 325). In the crude approximation of taking electron 1 to interact with a pointlike core of charge C_B , we would have $V_B = -C_B/r_{1B}$ and $\langle \mu_A | V_B | \nu_A \rangle$ would equal $-C_B \langle \mu_A | 1/r_{1B} | \nu_A \rangle$. Instead, similar to the CNDO/2 expression, MNDO takes $\langle \mu_A | V_B | \nu_A \rangle = -C_B(\mu_A \nu_A | s_B s_B)$, where s_B is a valence s orbital on atom B and the ERI $(\mu_A \nu_A | s_B s_B)$ is evaluated by an approximate procedure discussed briefly below. Thus

$$H_{\mu_A \nu_A}^{\text{core}} = -\sum_{B \neq A} C_B(\mu_A \nu_A | s_B s_B) \quad \text{for } \mu_A \neq \nu_A \quad (17.78)$$

The core matrix element $H_{\mu_A \mu_A}^{\text{core}} = \langle \mu_A(1) | \hat{H}^{\text{core}}(1) | \mu_A(1) \rangle$ containing the same AO twice is evaluated using (17.75) to write

$$H_{\mu_A \mu_A}^{\text{core}} = \langle \mu_A | -\frac{1}{2}\nabla^2 + V_A | \mu_A \rangle + \sum_{B \neq A} \langle \mu_A | V_B | \mu_A \rangle \quad (17.79)$$

The integral $U_{\mu_A \mu_A} \equiv \langle \mu_A | -\frac{1}{2}\nabla^2 + V_A | \mu_A \rangle$ could be evaluated from atomic spectral data (as it is in CNDO), but instead, MNDO takes this integral as a parameter. For an atom with s and p valence AOs, MNDO thus has parameters denoted U_{ss} and U_{pp} . MNDO evaluates the remaining integrals in $H_{\mu_A \mu_A}^{\text{core}}$ in (17.79) with the approximation (similar to that used in CNDO/2) $\langle \mu_A | V_B | \mu_A \rangle = -C_B(\mu_A \mu_A | s_B s_B)$, where the ERI is evaluated as discussed below. Thus

$$H_{\mu_A \mu_A}^{\text{core}} = U_{\mu_A \mu_A} - \sum_{B \neq A} C_B(\mu_A \mu_A | s_B s_B) \quad (17.80)$$

The ERIs are evaluated as follows in MNDO. The ZDO approximation makes all three-center and four-center ERIs vanish. The one-center ERIs are not evaluated by integration but are assigned values that give a good fit to valence-state ionization energies (which are known from atomic spectral data). The one-center ERIs are either Coulomb integrals of the form $g_{\mu\nu} \equiv (\mu_A \mu_A | \nu_A \nu_A)$ or exchange integrals of the form $h_{\mu\nu} \equiv (\mu_A \nu_A | \mu_A \nu_A)$. For an atom with only s and p valence AOs, there are six such one-center ERIs: $g_{ss}, g_{sp}, g_{pp}, g_{pp'}, h_{sp}, h_{pp'}$, where the prime denotes that the second p AO is along a different axis than the first. The values of these one-center ERIs found by fitting atomic data are less than the values given by direct integration, due to correlation effects that keep electrons apart.

The two-center ERIs are not evaluated directly by integration but are found from the values of the one-center ERIs and the internuclear distances using a complicated approximate procedure. [This procedure involves multipole expansions of the charge distributions; see M. J. S. Dewar and W. Thiel, *Theor. Chim. Acta*, **46**, 89 (1977).] The evaluation procedure gives values that are correct at the internuclear distances $R_{AB} = 0, R_{AB} = \infty$, and that are smaller in magnitude than the values found by accurate integration, so as to allow for electron correlation (and the use of a minimal basis set).

In MNDO, the core-core repulsion term is given by (17.76) with

$$f_{AB}^{\text{MNDO}} = C_A C_B (s_A s_A | s_B s_B) (e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}}) \quad (17.81)$$

where α_A and α_B are parameters for atoms A and B. For the pairs of atoms O–H and N–H, a different function than (17.81) is used; namely, the first term $e^{-\alpha_A R_{AH}}$ in parentheses in (17.81) is replaced with $(R_{AH}/\text{\AA})e^{-\alpha_A R_{AH}}$, where A is O or N.

In MNDO, there are thus six parameters to be optimized for each kind of atom: the one-center, one-electron integrals U_{ss} and U_{pp} , the STO orbital exponent ζ (MNDO takes $\zeta_s = \zeta_p$ for the valence AOs), β_s and β_p [Eq. (17.77)], and α [Eq. (17.81)]. For some

atoms, MNDO assumes $\beta_s = \beta_p$ and these atoms have five parameters. For H, there are no valence p orbitals and there are four parameters for H.

In 1985 Dewar and co-workers published an improved version of MNDO called **AM1** (Austin model 1, named for the University of Texas at Austin) [M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985)]. AM1 has been parametrized for nearly all the main-group elements and for Zn, Cd, and Hg. The only differences between MNDO and AM1 are that the AM1 valence orbital exponents ζ_s and ζ_p on the same atom are allowed to differ and the core-repulsion function in AM1 is given by (17.76) with

$$f_{AB}^{\text{AM1}} = f_{AB}^{\text{MNDO}} + \frac{C_A C_B}{R_{AB}/\text{\AA}} \left[\sum_k a_{kA} \exp[-b_{kA}(R_{AB} - c_{kA})^2] + \sum_k a_{kB} \exp[-b_{kB}(R_{AB} - c_{kB})^2] \right] \quad (17.82)$$

Here, f_{AB}^{MNDO} is given by (17.81); each sum in (17.82) contains two, three, or four Gaussian terms, depending which atom is involved; and the quantities a_{kA} , b_{kA} , and c_{kA} are parameters. [The expression given for f_{AB}^{AM1} in Dewar et al., *J. Am. Chem. Soc.*, **107**, 3902 (1985) is erroneous.] For example, for H there are three terms in the sum, giving nine additional parameters $a_{1H}, \dots, c_{3H}$. With this modified core-repulsion function, new optimized values for the original MNDO parameters and optimized values for the parameters in the sums in (17.82) were found.

In 1989, Stewart re-parametrized AM1 to give the **PM3 method** (parametric method 3, methods 1 and 2 being MNDO and AM1) [J. J. P. Stewart, *J. Comput. Chem.*, **10**, 209, 221 (1989); **11**, 543 (1990); **12**, 320 (1991)]. PM3 differs from AM1 as follows. The one-center electron-repulsion integrals are taken as parameters to be optimized (rather than being found from atomic spectral data). The core-repulsion function has only two Gaussian terms per atom. A different method was used to optimize the PM3 parameters. PM3 has been parametrized for nearly all the main-group elements and for Zn, Cd, and Hg.

The **RM1 method** (Recife Model 1, so named because it was developed at the Federal University of Pernambuco in Recife, Brazil) has exactly the same structure as AM1, but all 191 parameters for the atoms C, H, O, N, S, P, F, Cl, Br, I were reevaluated using data from 1736 molecules (as compared with about 200 molecules used for AM1) [G. B. Rocha et al., *J. Comput. Chem.*, **27**, 1101 (2006); www.rm1.sparkle.pro.br]. Since parameters are available for only 10 elements, RM1 is less widely applicable than AM1 or PM3.

The **PDDG/PM3** and **PDDG/MNDO** methods are modifications of PM3 and MNDO that add a certain function, called the pairwise distance directed Gaussian (PDDG) function, containing additional parameters, to the core-repulsion function, thereby significantly increasing the accuracy of these methods [M. P. Repasky et al., *J. Comput. Chem.*, **23**, 1601 (2002); I. Tubert-Brohman et al., *J. Comput. Chem.*, **25**, 138 (2004); *J. Chem. Theory Comput.*, **1**, 817 (2005)].

A major limitation of the original versions of the MNDO, AM1, and PM3 methods is that they use a basis set of s and p valence AOs only, so they cannot be used with transition-metal compounds. (In Zn, Cd, and Hg, the d electrons are not valence electrons.) Moreover, for compounds containing such second-row elements as S, the contributions of d orbitals to MOs are significant, and these methods do not perform well for such compounds.

Thiel and Voityuk [W. Thiel and A. A. Voityuk, *J. Am. Chem. Soc.*, **100**, 616 (1996)] extended MNDO to include d orbitals for many second-row and later elements, giving the **MNDO/d** method. MNDO/d does not add d orbitals for first-row elements, so for a compound containing only C, H, O, and N, MNDO/d is precisely the same as MNDO. Also, it was found that for Na, Mg, Zn, Cd, and Hg, inclusion of d orbitals made little difference,

and MNDO/d uses only an *sp* basis for these elements (but the MNDO parameters for these five elements must be reoptimized in MNDO/d). MNDO/d parameters have been published for Al, Si, P, S, Cl, Br, and I, and the method has also been parametrized for several transition metals.

Stewart revised the PM3 method to give the **PM5 method** (Stewart, MOPAC2002, Fujitsu Limited, Japan, 1999), which gives more accurate $\Delta H_{f,298}^\circ$ values than PM3 and has been parametrized for 50 elements, including many transition elements. (The PM4 method was never published, nor was it made available in a program. PM5 was never published.)

In 2007, Stewart published the **PM6 method** [J. J. P. Stewart, *J. Mol. Model.*, **13**, 1173 (2007)], which has been parametrized for 70 elements (nearly all the main-group elements and nearly all the transition elements). Data from 9000 compounds were used in the parametrization, including both experimental data and data from HF/6-31G* and B3LYP/6-31G* calculations. PM6 takes the core–repulsion term as (17.76) with

$$f_{AB} = C_A C_B (s_A s_A | s_B s_B) x_{AB} \exp[-\alpha_{AB}(R_{AB} + 0.0003R_{AB}^6)] + g_{AB}$$

where R_{AB} is in angstroms, $g_{AB} = 10^{-8}[(Z_A^{1/3} + Z_B^{1/3})/R_{AB}]^{12}$ (with Z_A and Z_B being the atomic numbers of A and B), and where x_{AB} and α_{AB} are two-atom parameters whose values depend on which atoms A and B are. The function g_{AB} is negligible at ordinary chemical distances, but becomes large in the rare situations where R_{AB} is very small; g_{AB} represents the repulsion between the cores of A and B. For the AB atom pairs OH and NH, the exponential in f_{AB} is replaced by $\exp(-\alpha_{AB}R_{AB}^2)$, so as to give a better representation of hydrogen bonding. Also, for the atom pair CC, an additional term is included in f_{AB} so as to improve the accuracy for compounds with carbon–carbon triple bonds.

With 70 elements, PM6 has $70 \cdot 69/2 + 70 = 2485$ different possible pairs of atoms for each of which pair a set of two-atom parameters is needed. Actually, PM6 includes parameters for about 450 pairs of atoms and covers the situations most commonly encountered. For a molecule with an atom pair not parametrized in PM6, the contribution of that pair to the core–core interaction is negligible if the two atoms are separated by at least 4 Å.

PM6 uses the real form of Slater orbitals as the basis functions. Of course, for transition elements, *d* orbitals are required in addition to *s* and *p* valence orbitals. To achieve improved performance, PM6 also includes *d* orbitals for many main-group nonmetals such as Si, P, S, Cl, As, Se, Br, Sb, Te, and I.

PM6 gives a significant improvement in accuracy over its predecessors. For example, for 1157 compounds containing no elements other than C, H, N, and O, the average absolute error in gas-phase $\Delta H_{f,298}^\circ$ values in kcal/mol is 9.4 for AM1, 5.7 for PM3, 5.6 for PM5, 4.9 for RM1, and 4.6 for PM6; for 1774 compounds containing no elements other than H, C, N, O, F, P, S, Cl, Br, I, these errors are 12.6 for AM1, 8.0 for PM3, 6.8 for PM5, 6.6 for RM1, and 5.1 for PM6 (Stewart, op. cit.); PM6 gives a better account of hydrogen bonding than its predecessors but is still deficient here.

Semiempirical methods usually give satisfactory bond lengths and bond angles, but their results are not as accurate as ab initio or DFT results with a suitable-size basis set. For 712 bond lengths and 244 bond angles in compounds containing no elements other than C, H, O, N, S, P, F, Cl, Br, and I, mean absolute errors in bond lengths in Å are 0.046 for AM1, 0.037 for PM3, 0.036 for RM1, and 0.031 for PM6; mean absolute errors in bond angles are 3.4° for AM1, 3.8° for PM3, 4.0° for RM1, and 3.2° for PM6 (Stewart, op. cit.). Semiempirical methods give rather mediocre results for torsional angles. For example, PM6 had a mean absolute error of 12.6° for 27 dihedral angles (Stewart, op. cit., supplementary material).

The semiempirical theories discussed so far do poorly in dealing with noncovalent interactions (hydrogen bonding, dispersion) that are key to determining the structures of biological molecules. The addition of parameter-containing empirical terms to the PM6

energy that correct for dispersion (D) and hydrogen bonding (H) gave the **PM6-DH2** method [M. Korth et al., *J. Chem. Theory Comput.*, **6**, 344 (2010)]. Further improvements in the dispersion and hydrogen-bonding correction terms gave the **PM6-DH+** and **PM6-D3H4** methods [M. Korth, *J. Chem. Theory Comput.*, **6**, 3808 (2010); J. Rezac and P. Hobza, *ibid.*, **8**, 141 (2012)]. These corrections for hydrogen bonding and dispersion have been parametrized for a few other semiempirical methods besides PM6.

The D3H4 energy correction is a function of the geometry of the molecule or interacting molecules. The energy in PM6-D3H4 has the form

$$E_{\text{PM6-D3H4}} = E_{\text{PM6}} + E_{\text{HB}} + E_{\text{disp}} + E_{\text{HHrep}}$$

where E_{HB} and E_{disp} are the corrections for hydrogen bonding and dispersion, and E_{HHrep} is a term representing repulsions between H atoms that was found necessary to achieve good results. In calculations of binding energies of hydrogen-bonded and dispersion-bonded complexes, PM6-D3H4 is found to be of similar accuracy to much more expensive DFT-D methods and is also pretty accurate in predicting geometries of such complexes. Averages of the root-mean-square errors in kcal/mol for eight test sets of intermolecular-interaction energies are as follows: 2.8 for B3LYP/6-31G*, 6.0 for AM1, 5.0 for PM3, 1.3 for PM3-D3H4, 1.1 for PM6-DH2, 1.1 for PM6-DH+, 0.85 for PM6-D3H4, and 0.81 for SCC-DFTB-D3H4 (discussed in the next subsection).

In 2012, Stewart published the **PM7 method** (openmopac.net/home.html; *J. Mol. Model.*, DOI: 10.1007/s00894-012-1667-x), a reparametrization of PM6 that improves the accuracy of heats of formation of organic compounds and that includes dispersion and hydrogen-bonding terms to better represent intermolecular interactions. PM7 has been parametrized for nearly all the main-group and transition elements. For a sample of 1366 compounds containing only C, H, O, N, F, Cl, S, P, and Br, mean absolute errors in gas-phase $\Delta H_{f,298}^{\circ}$ values in kcal/mol are 10.0 for AM1, 6.2 for PM3, 5.1 for B3LYP/6-31G*, 4.4 for PM6, and 4.0 for PM7. When compounds of all elements parametrized are included, PM7 gas-phase enthalpies of formation are substantially less accurate on average than those of PM6, a result likely due to inaccuracies in the reference data used to parametrize PM7. For the S66 test set of intermolecular interaction energies, average absolute errors in kcal/mol are 2.3 for B3LYP/6-31G*, 2.7 for PM6, 0.66 for PM6/DH2, 0.64 for PM6-DH+, and 0.78 for PM7.

Semiempirical methods are widely available in many programs. *Gaussian 09* (Section 15.14) includes the MNDO, AM1, PM3, PPDG, PM6, MINDO/3, INDO, and CNDO methods. Spartan includes the MNDO, MNDO/d, AM1, PM3, PM6, and RM1 methods. HyperChem has the MNDO, MNDO/d, AM1, PM3, RM1, MINDO/3, CNDO, INDO, INDO/S and extended Hückel methods. MOPAC2012 (openmopac.net/home.html) has the MNDO, AM1, PM3, PM6, PM6-DH2, PM6-DH+, PM7, and RM1 methods. It is available in Linux, Macintosh, and Windows versions and is free to academic researchers. An older version is MOPAC2009 (openmopac.net/MOPAC2009.html). AMPAC 9 (www.semichem.com) has the SAM1, AM1, PM3, RM1, PM6, MNDO, and MNDO/d methods and a molecule builder, and is available for Windows, Linux, and Macintosh operating systems.

The SCC-DFTB Method

The preceding methods discussed in this section are semiempirical MO methods. The SCC-DFTB method (self-consistent-charge density-functional tight-binding method) is a semiempirical DFT method [M. Elstner et al., *Phys. Rev. B*, **58**, 7260 (1998); M. Elstner, *Theor. Chem. Acc.*, **116**, 316 (2006)], somewhat similar to the semiempirical MO methods. The exchange-correlation energy functional used in SCC-DFTB is usually the PBE functional (Section 16.5). The SCC-DFTB method treats only the valence electrons

explicitly, uses a minimal basis set of nonorthogonal AOs to expand the Kohn–Sham orbitals, neglects many integrals, and uses approximations for many other integrals. The method contains two types of parameters: single-atom parameters (which are found from atomic DFT calculations), and parameters in interatomic repulsion-energy functions. The interatomic parameters are evaluated from bond-stretching energies calculated by DFT with the B3LYP functional with a DZP or TZP basis set. Thus the parameters are not found using experimental data, but using DFT calculations. The time for an SCC-DFTB calculation on a large molecule is similar to the time for an AM1 or PM3 calculation, but the results are usually more accurate than AM1 or PM3. The method is available in *Gaussian 09*. The D3H4 correction for hydrogen bonding and dispersion used in PM6-D3H4 has also been parametrized for SCC-DFTB, giving the DFTB-D3H4 method, whose accuracy for hydrogen-bonded and dispersion-bonded complexes is similar to that of PM6-D3H4 (Rezác and Hobza, op. cit.).

Semiempirical Calculations on Very Large Molecules

Methods that allow semiempirical calculations on molecules containing thousands of atoms are compared in A. D. Daniels and G. E. Scuseria, *J. Chem. Phys.*, **110**, 1321 (1999).

Yang and co-workers developed a “divide-and-conquer” method that combines calculations done on overlapping regions of the molecule [T.-S. Lee, D. M. York, and W. Yang, *J. Chem. Phys.*, **105**, 2744 (1996); S. L. Dixon and K. M. Merz, *J. Chem. Phys.*, **107**, 879 (1997)]. This method has been used for large biochemical systems.

Stewart developed the MOZYME linear-scaling method that uses localized MOs to solve the SCF equations for systems containing up to 15000 atoms [J. J. P. Stewart, *Int. J. Quantum Chem.*, **58**, 133 (1996)]. The method is limited to closed-shell molecules that can be represented by a conventional Lewis structure.

Stewart used MOZYME and PM6 to optimize the geometries of 45 proteins (including some containing metals) starting from the experimentally reported X-ray crystallography or NMR structure [J. J. P. Stewart, *J. Mol. Model.*, **15**, 765 (2009)]. The calculations were done on PCs having 1 or 2 GB of RAM. To speed up the computation, some approximations were made in evaluating certain integrals. The largest protein optimized had 14566 atoms. Each optimization finds the PM6 local minimum nearest the initial structure and does not find the global minimum. Stewart noted that because of the large computational time required, the practical size-limit of this method for routine work is about 9000 atoms. (Most of the proteins in the Protein Database contain less than 5000 atoms.) He suggested that such calculations could be useful in detecting large errors in experimentally determined protein structures and in refining such structures, and could be used to study enzyme–substrate reactions.

The *conjugate-gradient density-matrix search method* speeds up large-molecule SCF calculations by avoiding diagonalization of the Fock matrix [step 7 after Eq. (14.58)]; instead, an improved estimate of the density matrix $\mathbf{P}$ is found by varying the P_{tu} matrix elements so as to minimize the energy calculated from $\mathbf{P}$ and the current estimate of the Fock matrix $\mathbf{F}'$. Then the improved $\mathbf{P}$ is used to calculate an improved $\mathbf{F}'$, which is used to calculate an improved $\mathbf{P}$, and so on. This method allowed single-point AM1 energy calculations to be done on a 19995-atom polymer of glycine and a 6304-atom RNA molecule [A. D. Daniels, J. M. Millam, and G. E. Scuseria, *J. Chem. Phys.*, **107**, 425 (1997)].

The linear-scaling LocalSCF method [N. A. Anikin et al., *J. Chem. Phys.*, **121**, 1266 (2004)] uses weakly nonorthogonal localized MOs. A LocalSCF single-point AM1 calculation was done on a 119273-atom protein in 16000 s on a personal computer.

Semiempirical MO calculations on large systems are discussed in T. Clark and J. J. P. Stewart, “MNDO-like Semiempirical Molecular Orbital Theory”; Chap. 8 of J. R. Reimers (ed.), *Computational Methods for Large Systems*, Wiley, 2011. SCC-DFTB calculations on

such systems are discussed in M. Elstner and M. Gaus, “Self-Consistent-Charge Density Functional Tight-Binding Method,” chap. 9 of Reimers, op. cit.

17.5 The Molecular-Mechanics Method

The **molecular-mechanics (MM) method** is quite different from the semiempirical methods of the last section. Molecular mechanics is not a quantum-mechanical method, since it does not deal with an electronic Hamiltonian or wave function or an electron density. Instead, the method uses a model of a molecule as composed of atoms held together by bonds. Using such parameters as bond-stretching and bond-bending force constants, and allowing for interactions between nonbonded atoms, the method constructs a potential-energy expression that is a function of the atomic positions. By minimizing this expression for various molecular conformers, the MM method predicts equilibrium geometries and relative energies. The method was developed by Westheimer, Hendrickson, Wiberg, Allinger, Warshel, and others, and is applicable to ground electronic states. Most applications have been to organic compounds, but applications to organometallic compounds and transition-metal coordination compounds are growing. Because MM calculations are much faster than quantum-mechanical calculations, geometry optimizations of molecules with 10^4 atoms and single-point energy calculations on systems with 10^6 atoms can be done.

Molecular mechanics deals with the changes in a molecule's electronic energy due to bond stretching (V_{str}), bond-angle bending (V_{bend}), out-of-plane bending (V_{oop}), internal rotation (torsion) about bonds (V_{tors}), interactions between these kinds of motion (which produce cross terms V_{cross}), van der Waals attractions and repulsions between nonbonded atoms (V_{vdW}), and electrostatic interactions between atoms (V_{es}). The sum of these contributions gives the molecular-mechanics potential energy V , called the **steric energy**, for motion of the atoms in the molecule (or molecules if the system being calculated has more than one molecule):

$$V = V_{\text{str}} + V_{\text{bend}} + V_{\text{oop}} + V_{\text{tors}} + V_{\text{cross}} + V_{\text{vdW}} + V_{\text{es}} \quad (17.83)$$

Some people use the term *strain energy* to designate V in (17.83) but other people use strain energy to denote a different quantity (see *Burkert and Allinger*, pp. 184–189).

The explicit expressions used for each of the terms in (17.83) define what is called a molecular-mechanics **force field**, since the derivatives of the potential-energy function determine the forces on the atoms. A force field contains analytical formulas for the terms in (17.83) and values for all the parameters that occur in these formulas. The MM method is sometimes called the *empirical-force-field method*. Empirical force fields are used not only for single-molecule molecular-mechanics calculations of energy differences, geometries, and vibrational frequencies, but also for **molecular-dynamics** simulations of liquids and solutions, where Newton's second law is integrated to follow the motions of atoms with time in systems containing thousands of atoms.

An MM force field assigns each atom in a molecule to one of a number of possible **atom types**, depending on the atom's atomic number and molecular environment. For example, some commonly used atom types in force fields for organic compounds are sp^3 (saturated) carbon, sp^2 (doubly bonded) carbon, sp (triple bonded) carbon, carbonyl carbon, aromatic carbon, and so on, H bonded to C, H bonded to O, H bonded to N, and so on. Different force fields contain somewhat different numbers and kinds of atom types, based on the decisions made by their constructors. A force field for organic compounds typically contains 50 to 75 atom types.

In ab initio and semiempirical molecular electronic single-point or geometry-optimization calculations, one inputs the atomic numbers of the atoms and a set of coordinates (Cartesian or internal) for each atom, and no specification is made as to which

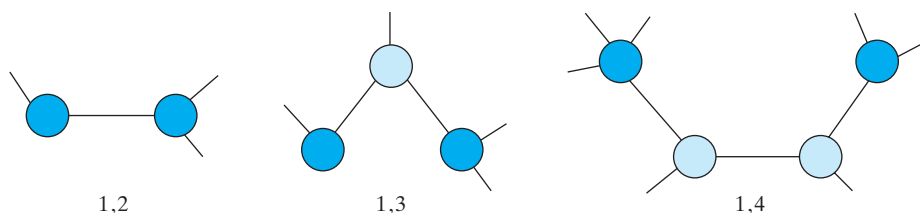


FIGURE 17.8 The heavily shaded atoms are 1,2 atoms, 1,3 atoms, and 1,4 atoms.

atoms are bonded to which atoms (Section 15.15). In a molecular-mechanics calculation, one must specify not only the initial atomic coordinates, but also which atoms are bonded to each atom, so that the V expression can be properly constructed. This specification is most conveniently done using a graphical computer interface to construct the molecule on screen (Section 15.15).

The following terminology (Fig. 17.8) is used in molecular mechanics: 1,2 atoms are atoms bonded to each other; 1,3 atoms are separated by two bonds; 1,4 atoms are separated by three bonds; and so on.

Some commonly used force fields are the following.

MM2 of Allinger and co-workers (*Burkert and Allinger*) is for small- to moderate-size organic compounds; MM3 is for small organic compounds, polypeptides, and proteins [N. L. Allinger and L. Yan, *J. Am. Chem. Soc.*, **115**, 11918 (1993) and references cited therein]; MM4 is an improved version [N. L. Allinger et al., *J. Comput. Chem.*, **24**, 1447 (2003); J. H. Lii et al., *ibid.*, **24**, 1473, 1490, 1504 (2003); C. H. Langley et al., *ibid.*, **22**, 1396, 1426, 1451, 1476 (2001)].

The following force fields are for polypeptides, proteins, and nucleic acids: AMBER (assisted model building with energy refinement) [W. D. Cornell et al., *J. Am. Chem. Soc.*, **117**, 5179 (1995); Y. Duan et al., *J. Comput. Chem.*, **24**, 1999 (2003); ambermd.org]; CHARMM (chemistry at Harvard molecular modeling) of Karplus and co-workers [B. R. Brooks et al., *J. Comput. Chem.*, **4**, 187 (1983); A. D. Mackerell et al., *J. Am. Chem. Soc.*, **117**, 11946 (1995); *J. Phys. Chem. B*, **102**, 3586 (1998); N. Foloppe and A. D. Mackerell, *J. Comput. Chem.*, **21**, 86 (2000); www.charmm.org]; OPLS (optimized potential for liquid simulations, sometimes called OPLS/AMBER) of Jorgensen and co-workers [W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1657 (1988); J. Pranata et al., *J. Am. Chem. Soc.*, **113**, 2810 (1991); OPLS-AA version—W. L. Jorgensen et al., *J. Am. Chem. Soc.*, **118**, 11225 (1996); G. A. Kaminski et al., *J. Phys. Chem. B*, **105**, 6474 (2001)]; and GROMOS (Groningen molecular simulation) of van Gunsteren and Berendsen [X. Daura, A. E. Mark, and W. F. van Gunsteren, *J. Comput. Chem.*, **19**, 535 (1998); C. Oostenbrink et al., *ibid.*, **25**, 1656 (2004); www.gromos.net]. CHARMM is a commercially available version of CHARMM that is also applicable to smaller organic compounds [F. A. Momany and R. Rone, *J. Comput. Chem.*, **13**, 888 (1992)]. MMFF94 (Merck molecular force field 1994 version) of Halgren [T. A. Halgren, *J. Comput. Chem.*, **17**, 490, 520, 553, 616 (1996); T. A. Halgren and R. B. Nachbar, *ibid.*, **17**, 587] is for calculations on small organic compounds, proteins, and nucleic acids. The Tripos force field (sometimes called SYBYL) [M. Clark et al., *J. Comput. Chem.*, **10**, 982 (1989)] is applicable to small organic molecules and proteins and is used for drug design.

CGenFF (CHARMM General Force Field) is similar to CHARMM but is specifically designed to be applied to drug-like molecules [K. Vanommeslaeghe et al., *J. Comput. Chem.*, **31**, 671 (2010); mackerell.umaryland.edu/~kenno/cgenff]. Unlike CHARMM, CGenFF is not applicable to biological macromolecules, but is meant to be used to represent ligands that might dock to a macromolecule such as a protein. The macromolecule

part of the system is modeled with CHARMM. (Pharmaceutical companies spend lots of effort looking for small molecules that might dock to the active sites of enzymes, and MM is the usual computational method in such studies.)

The DREIDING force field [S. L. Mayo, B. D. Olafson, and W. A. Goddard, *J. Phys. Chem.*, **94**, 8897 (1990)] is a general force field applicable to organic and inorganic compounds of all the main-group elements. The UFF (universal force field) [A. K. Rappé et al., *J. Am. Chem. Soc.*, **114**, 10024 (1992)] is applicable to compounds of all elements in the periodic table.

Most of these force fields exist in several versions, and one should specify the version used in any calculation.

Some force fields for MM calculations on biomolecules (for example, OPLS, AMBER, CHARMM) exist in both united-atom (UA) and all-atom (AA) versions. A UA force field saves computational time by not explicitly including hydrogen atoms bonded to aliphatic carbon atoms. Instead, the field contains parameters for the CH₃, CH₂, and CH groups.

For reviews of force fields for calculations on biomolecules, see A. D. Mackerell, *J. Comput. Chem.*, **25**, 1584 (2004); J. W. Ponder and D. A. Case, *Adv. Protein Chem.*, **66**, 27 (2003); W. L. Jorgensen and J. Tirado-Rives, *Proc. Nat. Acad. Sci.*, **102**, 6665 (2005).

One should distinguish between a force field, which is defined by the expression for V in (17.83), and a molecular-mechanics program, which is a computer program that uses a force field to perform molecular-mechanics calculations. Sometimes MM programs and force fields have the same name. For example, CHARMM is also the name of a molecular-mechanics program. The CHARMM program has available in it, not only the CHARMM force field, but also MMFF94.

Some examples of MM computational times on a workstation are: geometry optimization of a 29-atom molecule: 0.3 s for MM2* (a modified version of MM2), 80 s for MNDO, 2×10^5 s for HF/6-31G*; geometry optimization of a 182-atom molecule: 100 s for MM2*, 1.7×10^4 s for AM1 [C. H. Reynolds, *J. Mol. Struct. (Theochem)*, **401**, 267 (1997)].

We now consider the various terms in (17.83).

Stretching

The potential energy V_{str} of bond stretching is taken as the sum of potential energies $V_{\text{str},ij}$ for stretching each bond of the molecule: $V_{\text{str}} = \sum_{1,2} V_{\text{str},ij}$, where the sum is over all pairs of atoms bonded to each other, that is, over all 1,2 atom pairs. The simplest choice for $V_{\text{str},ij}$ is to use the harmonic-oscillator approximation (Fig. 4.6) and take $V_{\text{str},ij}$ as a quadratic function of the displacement of the bond length l_{ij} from its **reference (or natural) length** l_{IJ}^0 ; that is,

$$V_{\text{str},ij} = \frac{1}{2} k_{IJ} (l_{ij} - l_{IJ}^0)^2 \quad (17.84)$$

similar to (13.21). Here, the capital letters I and J denote the atom types of atoms i and j in the molecule. The force constant k_{IJ} and the reference length l_{IJ}^0 depend on the atom types of the two atoms forming the bond. For example, in CH₂=CH—CH₂—CH₃, each carbon–carbon single bond has a longer reference length and a smaller stretching force constant than the carbon–carbon double bond. Note that, because many terms in the steric energy (17.83) involve the atoms forming a particular bond, the reference value l_{IJ}^0 is not necessarily equal to the equilibrium bond length of the bond between atoms i and j in any particular molecule. However, it will likely be close to the equilibrium lengths of most bonds formed by atoms of types I and J . Many force fields omit the factor $\frac{1}{2}$ and write $V_{\text{str},ij} = K_{IJ} (l_{ij} - l_{IJ}^0)^2$.

The force fields AMBER, CHARMM, and Tripos use the quadratic function (17.84) for $V_{\text{str},ij}$. MM2 uses quadratic and cubic terms in $V_{\text{str},ij}$. MM3 and MMFF94 use quadratic, cubic, and quartic terms. For example, in MMFF94,

$$V_{\text{str},ij} = (143.93 \text{ kcal/mol}) \frac{1}{2} [k_{IJ}/(\text{mdyn}/\text{\AA})] (\Delta l_{ij}/\text{\AA})^2 [1 - (2 \text{ \AA}^{-1}) \Delta l_{ij} + \frac{7}{12} (2 \text{ \AA}^{-1})^2 \Delta l_{ij}^2] \quad (17.85)$$

where $\Delta l_{ij} \equiv l_{ij} - l_{IJ}^0$. It is traditional to use kcal/mol for molecular-mechanics energies, and the factor 143.93 converts from mdyn/Å to kcal/mol. The expression (17.85) has the form of the first three terms of a Taylor-series expansion of the Morse potential-energy function (Prob. 4.29) with the Morse exponential parameter equal to $2/\text{\AA}$; this value was found by trial and error to work well in MMFF94. The UFF and DREIDING force fields use a quadratic stretching function as the default but can also use Morse functions.

A particular force field will contain values for the parameters k_{IJ} and l_{IJ}^0 for all the possible bonds formed by the atom types of that force field. Because different force fields use different expressions for at least some of the terms in the steric energy (17.83), the parameters of one force field are not really comparable to those of another field. (A mistake people sometimes make is to supply a missing parameter in one force field by using a parameter from another force field.) Some examples of l_{IJ}^0 values in Å are: for a bond between two sp^3 carbons, 1.508 in MMFF94, 1.54 in Tripos 5.2, 1.523 in MM2(87); for a bond between an sp^3 carbon and an sp^2 carbon, 1.482 in MMFF94, 1.501 in Tripos 5.2, 1.497 in MM2(87); for a double bond between two sp^2 carbons, 1.333 in MMFF94, 1.335 in Tripos 5.2, 1.337 in MM2(87).

Bending

The potential energy V_{bend} of bond bending is taken as the sum of potential energies $V_{\text{bend},ijk}$ for bending each bond angle of the molecule: $V_{\text{bend}} = \sum V_{\text{bend},ijk}$, where the sum is over all bond angles in the molecule. The simplest choice is to take $V_{\text{bend},ijk}$ as a quadratic function:

$$V_{\text{bend},ijk} = \frac{1}{2} k_{IJK} (\theta_{ijk} - \theta_{IJK}^0)^2 \quad (17.86)$$

where θ_{IJK}^0 is the reference value for the bond angle type IJK . This form is used in CHARMM, AMBER, and Tripos. MMFF94 uses quadratic and cubic terms in $V_{\text{bend},ijk}$ and uses a special form for bond angles near 180° . MM2 uses quadratic and sextic terms.

Torsion

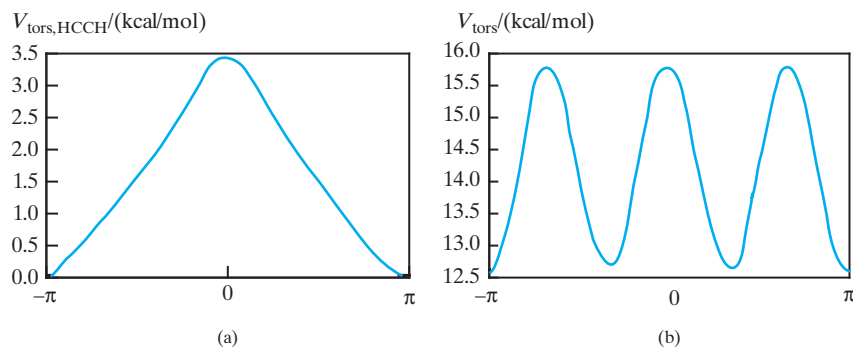
The term V_{tors} is taken as the sum of terms $V_{\text{tors},ijkl}$ over all 1,4 atom pairs: $V_{\text{tors}} = \sum_{1,4} V_{\text{tors},ijkl}$. For example, for ethane, H_3CCH_3 , each hydrogen on the left carbon has a 1,4 relation to each of the three hydrogens on the right carbon, giving a total of nine terms in the V_{tors} sum. For H_2CCH_2 , there are four torsion terms in the sum. A commonly used form (MMFF94, MM2, MM3) for $V_{\text{tors},ijkl}$ is

$$V_{\text{tors},ijkl} = \frac{1}{2} [V_1(1 + \cos \phi) + V_2(1 - \cos 2\phi) + V_3(1 + \cos 3\phi)] \quad (17.87)$$

where ϕ is the dihedral angle $D(ijkl)$ (Fig. 15.14) and V_1, V_2, V_3 are parameters whose values depend on the atom types of i, j, k , and l . Tripos, CHARMM, and DREIDING use the following simpler form containing only one parameter: $\frac{1}{2} V_n [1 + \cos(n\phi - \phi_0)]$, where n gives the number of minima over 360° of the torsional potential and ϕ_0 determines the locations of the minima. For example, for $n = 3$ and $\phi_0 = 0$, we get a potential with minima at $60^\circ, 180^\circ$, and 300° , as in C_2H_6 ; with $n = 2$ and $\phi_0 = \pi$, we get a potential with minima at 0° and 180° , as in C_2H_4 .

The potential-energy change as one methyl group is rotated relative to the other in ethane can be determined experimentally by observation of transitions between vibrational energy levels produced by this potential. In an MM force field, V_{tors} in ethane is the sum of contributions $V_{\text{tors},ijkl}$ from nine dihedral angles, and an individual $V_{\text{tors},ijkl}$ is not experimentally determinable. By taking the appropriate second derivative of the electronic energy of ethane and integrating the result, one can use ab initio calculations to estimate

FIGURE 17.9 (a) HCCH torsional potential in ethane found from HF/6-31G* calculations with bond distances and bond angles held fixed during torsion [U. Dinur and A. T. Hagler, *J. Comput. Chem.*, **11**, 1234 (1990)]. (b) Sum of the nine HCCH torsional potentials of (a).



$V_{\text{tors},ijkl}$. The result from a HF/6-31G* calculation is shown in Fig. 17.9a. When expressed as a series of terms involving $\cos n\phi$, this curve turns out to have significant contributions from only the $\cos \phi$, $\cos 2\phi$, and $\cos 3\phi$ terms. When the nine dihedral-angle contributions are added, the symmetry of C_2H_6 makes the contributions of the $\cos \phi$ and $\cos 2\phi$ terms add to zero, leaving only the $\cos 3\phi$ term (Fig. 17.9b). Because of insufficient experimental data, MM2 and MM3 take V_1 and V_2 as zero for HCCH torsion (where C is a saturated carbon), but MMFF94 used results of ab initio conformational energy calculations to fit nonzero values for these quantities.

By using a large positive value for V_2 in (17.87), one can produce the high barrier to out-of-plane twisting in $\text{H}_2\text{C}=\text{CH}_2$.

Out-of-Plane Bending

The molecule cyclobutanone, $\text{C}_3\text{H}_6\text{C}=\text{O}$, contains an oxygen doubly bonded to one of the carbons of a four-membered ring. The carbonyl carbon and the three atoms bonded to it all lie in the same plane. The CCC bond angle at the carbonyl carbon is close to 90° , and the CCO bond angles are close to $\frac{1}{2}(360^\circ - 90^\circ) = 135^\circ$. Because these 135° angles deviate considerably from the reference angles of approximately 120° at the carbonyl carbon, the bond-bending terms $V_{\text{bend},ijk}$ at the carbonyl carbon will make the equilibrium position of the oxygen atom lie above the plane of the ring carbons, so as to produce near- 120° CCO angles. To ensure planarity of the three atoms bonded to a carbonyl carbon (which is favored by pi bonding between C and O), one includes in V an out-of-plane (oop) bending term at each carbonyl carbon. This term can have the form $\frac{1}{2}k_{\text{oop}}\chi_{\text{oop}}^2$, where χ_{oop} is the angle between the CO bond and the plane of the carbonyl carbon and the two carbons bonded to it. (In actual practice, one either includes an out-of-plane bending term for each of the three atoms bonded to a carbonyl C or uses a single term where χ is taken as the average of the out-of-plane angles for each of the three atoms bonded to the carbonyl C.) Similar out-of-plane bending terms are used to enforce planarity at nitrogen atoms in amides. [The question of planarity of the amide group is complicated; see T. A. Halgren, *J. Comput. Chem.*, **17**, 553 (1996); G. Forgarasi and P. G. Szalay, *J. Phys. Chem. A*, **101**, 1400 (1997).] Also, an out-of-plane bending term is often used at a carbon double bonded to another carbon, since this allows vibrational frequencies to be better reproduced.

Instead of out-of-plane bending terms, some force fields use so-called improper torsion terms to achieve the same result.

Cross Terms

The cross terms V_{cross} in (17.83) allow for interactions between stretching, bending, and torsional motions. For example, if the CO and OH bonds of a COH bond angle are stretched,

then the distance between the end atoms of the COH bond angle is increased, which makes it easier to bend the COH angle. Likewise, reducing the COH angle tends to increase the OH and CO lengths. To allow for this interaction, one can add a stretch–bend cross term with the form $\frac{1}{2}k_{12}(\Delta l_1 + \Delta l_2)\Delta\theta$, where Δl_1 , Δl_2 , and $\Delta\theta$ are the deviations of the bond lengths and the bond angle from their reference values.

The most commonly used cross terms are stretch–bend, stretch–stretch for two bonds to the same atom, stretch–torsion, bend–torsion, and bend–bend for two angles with a common central atom. MM2 and MMFF94 include only stretch–bend interactions. The TRIPOS, AMBER, CHARMM, DREIDING, and UFF force fields have no cross terms. MM3 and MM4 include stretch–bend, bend–bend, and stretch–torsion terms.

Electrostatic Interactions

The electrostatic term V_{es} is commonly taken as the sum of electrostatic interactions involving all pairs of atoms except 1,2 and 1,3 pairs: $V_{\text{es}} = \sum_{1 \leq i < j \leq 4} V_{\text{es},ij}$, where atoms i and j have a 1,4 or greater relation. $V_{\text{es},ij}$ is calculated by assigning partial atomic charges Q_i to each atom and using the Coulombic potential energy expression

$$V_{\text{es},ij} = \frac{Q_i Q_j}{\epsilon_r 4\pi\epsilon_0 R_{ij}} \quad (17.88)$$

where ϵ_r is a dielectric constant. For calculations modeling gas-phase molecules, ϵ_r is typically given a value in the range 1 to 1.5. An ϵ_r greater than 1 allows for screening interactions due to polarization of parts of the molecule lying between atoms i and j .

Various methods are used to assign the partial atomic charges. MMFF94 assumes the polarity of a given type of bond to be independent of its environment and assigns charges according to $Q_i = Q_{i,\text{formal}} + \sum \omega_{\text{KI}}$, where $Q_{i,\text{formal}}$ is the formal charge on atom i (found from the Lewis dot structure by dividing the electrons in each bond equally between the bonded atoms), ω_{KI} is a parameter representing the charge contribution to atom i from the bond between atoms i and k (of types I and K), and the sum goes over all bonds to atom i . Note that $\omega_{\text{KI}} = -\omega_{\text{IK}}$. Some values of ω_{KI} are 0 for a $\text{Csp}^3\text{--H}$ bond and -0.138 for a $\text{Csp}^3\text{--Csp}^2$ bond. The ω_{KI} values were found by varying them to make MMFF94 give a good least-squares fit to HF/6-31G* dipole moments of several hundred molecules; the value zero for $\text{Csp}^3\text{--H}$ bonds was assumed. [It should be noted that MMFF94 replaces R_{ij} in (17.88) with $R_{ij} + 0.05 \text{ \AA}$.]

In AMBER, the charges are found by using charges fit to HF/6-31G* electrostatic potentials (ESPs, Section 15.7) of related smaller molecules. For example, to get the charges on atoms of the amino acid residue --NHCH(R)C(O)-- of a protein, one uses the charges of corresponding atoms of the molecule $\text{CH}_3\text{C(O)NHCH(R)C(O)NH}_2$. Similarly for the charges in DNA and RNA.

In OPLS, partial charges are assigned to the atoms in a protein based on the atom type. For example, all oxygen atoms in amide groups are given a partial charge of -0.50 and all carbons in amide groups are given a charge of $+0.50$.

Several other schemes based on electronegativity are used to assign charges; see *Leach*, Section 4.9.6; T. A. Halgren, *J. Comput. Chem.*, **17**, 520 (1996).

MM2 and MM3 do not use (17.88) but assign a dipole-moment vector to each bond and compute V_{es} as the sum of the potential energies of interaction between bond moments. Each bond dipole moment is located at the center of the bond and points along the bond. The values for the bond moment of given types of bond were chosen to fit known dipole moments of small molecules.

Some force fields reduce $V_{\text{es},ij}$ of each 1,4 pair of atoms by multiplying it by a scaling factor. Some 1,4 electrostatic scaling factors are 0.75 in MMFF94 and 0.5 in AMBER.

Van der Waals Interactions

The van der Waals term in (17.83) is usually taken as the sum of interactions involving all possible 1,4, 1,5, 1,6, . . . atom-pair interactions: $V_{\text{vdW}} = \sum_{1 \leq i < j} V_{\text{vdW},ij}$, where atoms i and j are in a 1,4 or greater relation. The 1,2 and 1,3 van der Waals interactions and electrostatic interactions are considered to be implicitly included in the bond-stretching and bond-bending parameters. Each van der Waals pair term $V_{\text{vdW},ij}$ is the sum of an attraction due to London dispersion forces and a repulsion due mainly to Pauli repulsion. The AMBER, CHARMM, DREIDING, UFF, and TRIPOS force fields take $V_{\text{vdW},ij}$ as a Lennard-Jones 12–6 potential (Fig. 17.10 and Prob. 17.38), which can be written in two equivalent forms:

$$V_{\text{vdW},ij} = \epsilon_{IJ} \left[\left(\frac{R_{IJ}^*}{R_{ij}} \right)^{12} - 2 \left(\frac{R_{IJ}^*}{R_{ij}} \right)^6 \right] = 4\epsilon_{IJ} \left[\left(\frac{\sigma_{IJ}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{IJ}}{R_{ij}} \right)^6 \right] \quad (17.89)$$

where R_{ij} is the distance between atoms i and j , the well-depth parameter ϵ_{IJ} is the value of $V_{\text{vdW},ij}$ at the minimum in the interaction curve, the parameter R_{IJ}^* gives the value of R_{ij} at the minimum in $V_{\text{vdW},ij}$ and the alternative parameter σ_{IJ} is the R_{ij} value at which $V_{\text{vdW},ij}$ is zero. MM2 and MM3 use a function with an attractive term proportional to R_{ij}^{-6} and a repulsive term proportional to $e^{-aR_{ij}}$, where a is a parameter. MMFF94 uses a buffered 14–7 potential that was found to represent known van der Waals interactions between the rare gas atoms more accurately than other forms [T. A. Halgren, *J. Am. Chem. Soc.*, **114**, 7827 (1992)].

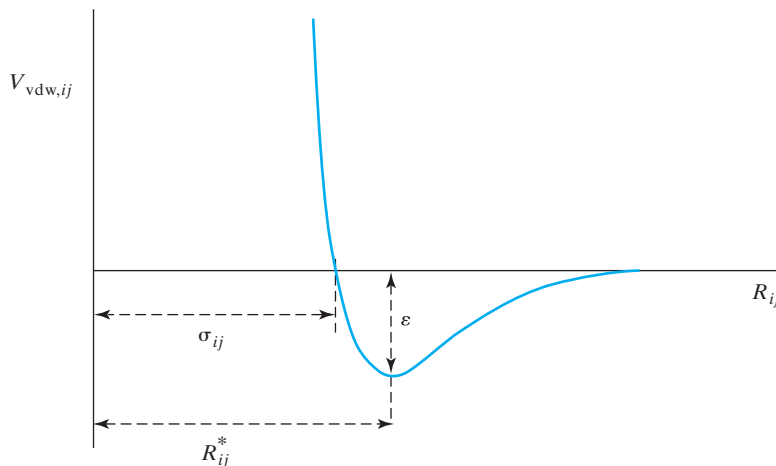
For 1,4 interactions, some force fields multiply $V_{\text{vdW},ij}$ by a scale factor that reduces its value.

As an example, for $\text{CF}_3\text{CF}_2\text{CF}_3$, a molecular-mechanics force field has 10 bond-stretching terms (8 for CF bonds and 2 for CC bonds), 18 bond-bending terms [there are $\frac{1}{2}(4)3 = 6$ bond angles at each C], 18 bond torsion terms (9 about each CC bond), 27 van der Waals terms (there are 18 1,4 interactions between the end atoms of the 18 dihedral angles plus 9 1,5 interactions between one set of CF_3 fluorines and the second set), 27 electrostatic terms, plus whatever cross terms the force field includes.

Cutoffs

The electrostatic and van der Waals interactions are called **nonbonded interactions** and consume the largest part of the time needed to calculate V_{steric} of a very large molecule. For a 3000-atom molecule, V_{es} and V_{vdW} are each the sum of about $\frac{1}{2}(3000)(2990) \approx 4 \times 10^6$ terms. (The last factor is reduced to acknowledge that 1,2 and 1,3 interactions are

FIGURE 17.10 The Lennard-Jones potential as a function of interatomic distance.



omitted.) To speed up MM calculations on large molecules (and molecular dynamics calculations on systems containing many molecules), many programs use a **cutoff**; that is, $V_{es,ij}$ and $V_{vdW,ij}$ terms are omitted for atom pairs that are farther apart than some chosen distance.

If a cutoff is abruptly applied at a particular interatomic distance, this discontinuity can cause problems in energy minimization and molecular-dynamics calculations. To avoid this, one can use a cutoff that makes the nonbonded interactions go to zero gradually over a distance of, say, 1 Å. This is done using what is called a *switching function*.

Van der Waals interactions are proportional to $1/R_{ij}^6$ at large distances and so are short range. A van der Waals cutoff distance of 8 or 10 Å is typically used and was believed to produce little error. However, molecular-dynamics simulations of liquid alkanes using the GROMOS force field found that the enthalpies of vaporization and the vapor pressures changed very substantially when the van der Waals cutoff distance was varied within the range 8 to 14 Å. Thus, a van der Waals cutoff radius of 16 Å is much more justifiable than the traditional one of 8 or 10 Å [X. Daura et al., *J. Comput. Chem.*, **19**, 535 (1998)].

Often an electrostatic cutoff distance of 10 or 15 Å is used. However, electrostatic interactions are long range; thus use of a cutoff distance for them is not justifiable and produces serious errors. Instead of using a cutoff, one can calculate the electrostatic interactions using the fast-multipole method (FMM; Section 15.16), but this method may not be faster than direct summation of all the pairwise interactions when the system has less than 3000 atoms. The *structure-adapted multipole method* (SAMM) [C. Niedermeier and P. Tavan, *J. Chem. Phys.*, **101**, 734 (1994)] achieves faster evaluation of the electrostatic energy than the FMM at a small sacrifice in accuracy. In SAMM, instead of using the fixed subdivision of space that is used in the FMM method, space is divided into regions that correspond to various structural features of the biomolecule being studied, thereby allowing the multipole expansions to be truncated at lower orders than in the FMM method. For molecular-dynamics calculations on biomolecules, the SAMM method has been combined with a multiple-time-step method to yield FAMUSAMM (fast multiple-time-step structure-adapted multipole method); M. Eichinger et al., *J. Comput. Chem.*, **18**, 1729 (1997). In a multiple-time-step method, the forces between distant atoms are treated using a longer time step than is used for the forces between nearby atoms (which vary more rapidly).

The most widely used method to sum the electrostatic interactions in large systems is the *particle-mesh Ewald method*, which is a modification of a method originally used to sum electrostatic interactions in ionic crystals. Here, the electrostatic-energy sum is replaced by the sum of sums that can each be calculated more rapidly than the original sum; see *Leach*, Section 6.8.1.

Hydrogen Bonding

For pairs of atoms that can hydrogen bond to each other, some force fields modify the van der Waals interaction to a form such as $A/R^{12} - C/R^{10}$, but many force fields contain no special terms for hydrogen bonding, and rely on the electrostatic and van der Waals terms to produce the hydrogen bond.

Conjugated Bonds

Conjugated bonds require special consideration in molecular mechanics. Consider, for example, $\text{CH}_2=\text{CH}-\text{CH}=\text{CH}_2$. Each of the four carbons is an sp^2 type. Hence, if the stretching force constant and reference length of each bond is assigned simply based on the types of the two atoms forming the bond, each carbon-carbon bond in 1,3-butadiene will have the same stretching constant and reference length. However, the central and end bonds actually have quite different lengths (1.46 and 1.34 Å). MMFF94 and Tripos make the reference bond lengths and stretching force constants depend not only on the atom types but also on the bond order in the Lewis structure. For two bonded sp^2 carbons,

the MMFF94 reference length is 1.333 Å if the bond is a single bond and is 1.430 Å if it is a double bond, and the stretching constant for the double bond is about twice that for the single bond. Some force fields handle this problem in a slightly different way that is equivalent to the MMFF94 procedure. MM2, MM3, and MM4 use the more general procedure of performing a semiempirical π -electron calculation on the conjugated portion of the molecule to derive bond orders and use these orders to assign reference bond lengths and force constants to the conjugated bonds.

Parametrization

The general procedure for finding parameter values in a force field is to use experimental or ab initio theoretical information to choose an initial set of parameter values and then vary these parameters so as to minimize the deviations of force-field predicted molecular properties from experimental or ab initio calculated properties of a chosen set of molecules, called the **training set**. Molecular properties used in force-field parametrizations include molecular structures, conformational energy differences, vibrational frequencies, barriers to internal rotation, dipole moments, and intermolecular interactions. Earlier force fields relied mainly on experimental data in the parametrization process, but as ab initio calculations become increasingly able to accurately treat larger molecules, ab initio calculated properties are increasingly being used in parametrization (examples are MMFF94 and CGenFF).

Initial estimates of the reference length l_{IJ}^0 and angle θ_{IJK}^0 can be taken as the typical experimental or ab initio calculated length of an IJ bond and an IJK angle. These initial values can be refined by adjusting them to minimize errors in predicted bond lengths and angles of the training-set molecules.

Bond-bending and stretching force constants in polyatomic molecules are not experimentally observable quantities, since the observed vibrational frequencies are related in a complicated way to such force constants. However, one can write an approximate expression for the molecular vibrational potential energy as a sum of quadratic terms involving bond-stretching, bending of independent bond angles, torsion involving independent dihedral angles, plus various cross terms. (Such a *valence force field* is not the same as a molecular-mechanics force field. In an MM force field, the bond-bending angles and the dihedral angles are not all independent of one another, and the valence force field omits nonbonded interactions.) By using a valence force field and combining vibrational-frequency data of many unstrained organic compounds, Schachtschneider and Snyder obtained bond-stretching and bending force constants that are approximately transferable (*Burkert and Allinger*, pp. 20–21), and these can be used as initial estimates for these parameters. An alternative way to get initial estimates of the force constants is from analytical second derivatives of ab initio molecular electronic energies. The initial force constants can be improved by adjusting their values to minimize the errors in vibrational frequencies predicted by the force field.

An initial estimate of V_3 for HCCH (where C is saturated) can be found by fitting the rotational barrier in C_2H_6 . Then V_3 for CCCH can be found by fitting the barrier in $CH_3CH_2CH_3$ and V_3 for CCCC from $CH_3CH_2CH_2CH_3$ (with V_1 and V_2 assumed to be zero). To refine V_3 and find V_1 and V_2 values, one can fit conformational energy differences (experimental or ab initio) for a set of training molecules. V_2 for $HC=CH$ can be fit to the rotational barrier in C_2H_4 .

For van der Waals parameters, one usually assumes that the parameters ϵ_{IJ} and R_{IJ}^* for interaction between unlike atom types can be calculated from the parameters ϵ_{II} , ϵ_{JJ} , R_{II}^* , R_{JJ}^* for like interactions using mixing rules. The most common mixing rules are $\epsilon_{IJ} = (\epsilon_{II}\epsilon_{JJ})^{1/2}$ and $R_{IJ}^* = \frac{1}{2}(R_{II}^* + R_{JJ}^*)$ or $R_{IJ}^* = (R_{II}^*R_{JJ}^*)^{1/2}$. MMFF94 uses more elaborate mixing rules, which were found by examining known rare-gas interaction data.

Sublimation energies of molecular compounds depend on van der Waals interactions, and such sublimation-energy data have been used to derive values of van der Waals parameters. Some force fields (for example, AMBER94, OPLS) adjust van der Waals parameters to give good values for properties such as densities and enthalpies of vaporization found in Monte Carlo simulations of liquid compounds.

Since all the parameters in a force field will influence all calculated molecular properties, parametrization is an iterative process. After adjusting a particular subset of parameters to fit some molecular property, the previously done fits of parameter subsets will be worsened, so one goes back and readjusts previously fitted parameters. One keeps iterating until no further significant improvement is obtained.

To use the molecular-mechanics method, one needs enough data to parametrize the force field. This makes it hard to apply molecular mechanics to novel types of compounds, but *ab initio* calculations may help in the parametrization.

An empirical force field may contain hundreds or thousands of parameters. MMFF94 has roughly the following numbers of parameters: 500 stretching force constants, 500 reference bond lengths, 2300 bending force constants, 2300 reference bond angles, 600 stretch-bend constants, 100 out-of-plane force constants, 2800 torsion parameters of which roughly half are zero, leaving about 1400 nonzero ones, 400 van der Waals parameters, and 600 electrostatic charge-increment parameters, for a total of about 9000 parameters. In contrast, UFF, which has 126 atom types and can treat compounds of all elements, contains only about 800 parameters. The number of parameters is kept relatively small in UFF by devices such as taking the reference bond angles θ_{JJK}^0 to depend only on the atom type of the central atom J, calculating reference bond lengths l_{IJ}^0 as the sum of bond radii of the atom types plus corrections for bond order and for electronegativity differences, calculating stretching force constants by an empirical rule that takes them as functions of l_{IJ}^0 and effective-charge parameters for each atom type, and so on. Because of the relatively small number of parameters, UFF cannot achieve the accuracy of highly parametrized force fields such as MM3 or MMFF94, but the loss of accuracy is a trade-off for the broad applicability of UFF.

Molecular Properties

What molecular properties can be calculated using molecular mechanics? A molecular-mechanics geometry optimization starts with the initially assumed geometry and finds the nearest local energy minimum by minimizing the steric energy V of (17.83) using one of the methods of Section 15.10. Because (17.83) provides an analytical expression for the energy, the first and second derivatives of V can be easily evaluated analytically, which facilitates energy minimization. Minimization of V yields the MM-predicted equilibrium geometry of a particular conformer. Many molecular-mechanics programs have built-in searching methods (Section 15.11) that locate many low-energy conformers.

One very practical use of molecular mechanics is to refine the structure produced by a molecule builder (Sec. 15.15). After one draws onscreen the molecule to be studied by an *ab initio*, density functional, or semiempirical method, molecule builders in most programs allow you to refine the structure by doing a quick molecular mechanics calculation to locate the local MM energy minimum nearest the structure drawn. By starting from the MM structure, the geometry optimization will take less time.

It must be emphasized that the numerical value of the equilibrium steric energy of a conformer has no physical significance by itself. The zero level of V corresponds to a fictitious molecule in which all the bond lengths and angles have their reference values and torsional, van der Waals, and electrostatic interactions are absent. Different force fields will typically give very similar equilibrium geometries for a given conformer, but will give quite different steric energies for that conformer. Steric energies depend on how the force field was constructed and parametrized. In contrast, the electronic energy including

nuclear repulsion found by an ab initio calculation has the physical significance of being the approximate energy relative to a well-defined system with all electrons and nuclei infinitely far apart from one another.

What does have physical significance in molecular mechanics is the steric-energy difference (calculated using the same force field) between two species having the same numbers and kinds of atoms and the same numbers and kinds of bonds. Thus, one can use differences in steric energies to meaningfully calculate energy differences between: (a) different conformers of the same molecule [for example, the anti and gauche conformers of butane (Section 15.10) or the chair and boat conformers of cyclohexane]; (b) different stereoisomers of the molecule (for example, cis and trans 1,2-dichloroethylene); (c) species differing by rotation about a bond (for example, eclipsed and staggered ethane); (d) different geometries of the same molecule (for example, planar NH_3 and pyramidal NH_3); (e) two molecules far apart from each other and the same pair forming a hydrogen bond. Intermolecular-interaction energies can be calculated.

The difference between the molecular-mechanics equilibrium steric energies of two species with the same numbers and kinds of bonds is an estimate of the difference between the equilibrium electronic energies of the two species. Corrections for zero-point vibrational energy differences and thermal energy differences should be added to the steric energy difference, but these are usually small and are often omitted.

A molecular-mechanics steric energy can be combined with empirical bond-energy parameters to calculate gas-phase heats of formation $\Delta H_{f,298}^\circ$, as discussed later in this section.

Although the empirical force fields discussed in this section provide a good representation of potential-energy surfaces in the regions of minima, one cannot use them to calculate the complete potential-energy surface for a chemical reaction, since these fields are incapable of describing bond breaking.

By taking $\sum_i Q_i r_i$ [Eq. (14.9)], where the sum goes over all the partial atomic charges, one can calculate a molecular dipole moment for a given conformation.

By evaluating the second partial derivatives of V at a local minimum, one can use the procedure of Section 15.12 to calculate molecular vibrational frequencies. Using the calculated structure and vibrational frequencies, one can estimate the gas-phase entropy S_{298}° (Section 15.13).

Molecular-mechanics force fields are widely used in molecular dynamics simulations that integrate the motions of atoms over time. For example, a remarkable series of all-atom simulations of the opening and closing of a voltage-gated cell-membrane potassium-ion channel used versions of the CHARMM force field; these simulations used from 100000 to 230000 atoms followed for 150 to 250 μs [M. Jensen et al., *Science*, **336**, 229 (2012)] and were done on a special-purpose parallel computer [en.wikipedia.org/wiki/Anton_(computer)]. Movies of the simulations are available at www.sciencemag.org/content/336/6078/229/suppl/DC1. (Several protein-folding simulations can be viewed at YouTube.com.)

Heats of Formation

The molecular-mechanics zero level of energy has all bond lengths and angles at their reference values and no electrostatic, van der Waals, or torsional interactions. In such a hypothetical state, one can well approximate the molecular binding energy as the sum of empirical bond energies. Therefore, the equilibrium electronic energy U_{eq} of a molecule can be found by combining the molecular-mechanics-calculated equilibrium-geometry steric energy V_{steric} with bond energies. For example, for a saturated hydrocarbon with formula $\text{C}_{n_{\text{C}}}\text{H}_{n_{\text{H}}}$ (where n_{C} and n_{H} are the number of C atoms and H atoms), we have $U_{\text{eq}} = V_{\text{steric}} - n_{\text{CH}}b_{\text{CH}} - n_{\text{CC}}b_{\text{CC}}$, where n_{CH} and n_{CC} are the numbers of CH and CC

single bonds in the molecule, b_{CH} and b_{CC} are CH and CC bond energies (which by convention are positive), and the zero level of energy is taken to correspond to separated (gas-phase) atoms $\text{C}(g)$ and $\text{H}(g)$. In the following discussion, all energies will be on a per mole basis.

For the formation reaction



the change in molar equilibrium electronic energy is

$$V_{\text{steric}} - n_{\text{CH}}b_{\text{CH}} - n_{\text{CC}}b_{\text{CC}} - n_{\text{C}}U_{\text{C}} - \frac{1}{2}n_{\text{H}}U_{\text{H}_2}$$

where U_{C} and U_{H_2} are the molar equilibrium electronic energies of graphite and $\text{H}_2(g)$ with respect to the same zero level of energy as above. If we temporarily ignore zero-point vibrational energy, we can take this change in equilibrium electronic energy as equal to the change in standard-state thermodynamic internal energy for the formation reaction at absolute zero:

$$\Delta U_{f,0}^{\circ} = V_{\text{steric}} - n_{\text{CH}}b_{\text{CH}} - n_{\text{CC}}b_{\text{CC}} - n_{\text{C}}U_{\text{C}} - \frac{1}{2}n_{\text{H}}U_{\text{H}_2}$$

For a saturated hydrocarbon (acyclic or cyclic), it is not hard to see that the numbers of H and C atoms are related to the numbers of CH and CC bonds by

$$n_{\text{H}} = n_{\text{CH}} \quad \text{and} \quad n_{\text{C}} = \frac{1}{4}n_{\text{CH}} + \frac{1}{2}n_{\text{CC}}$$

At temperature T , the change in translational energy for the above formation reaction is $\frac{3}{2}RT - \frac{1}{2}n_{\text{H}}(\frac{3}{2}RT)$, and the change in rotational energy is $\frac{3}{2}RT - \frac{1}{2}n_{\text{H}}RT$. The relation $\Delta H_f^{\circ} = \Delta U_f^{\circ} + \Delta(PV)^{\circ} = \Delta U_f^{\circ} + \Delta n_g RT$, where Δn_g is the change in number of moles of gas in the formation reaction, gives $\Delta H_f^{\circ} = \Delta U_f^{\circ} + (1 - \frac{1}{2}n_{\text{H}})RT$. Combining all these relations and continuing to ignore the contribution from the change in vibrational energy, we have for the standard enthalpy of formation at temperature T

$$\Delta H_{f,T}^{\circ} = V_{\text{steric}} - n_{\text{CH}}b_{\text{CH}} - n_{\text{CC}}b_{\text{CC}} - (\frac{1}{4}n_{\text{CH}} + \frac{1}{2}n_{\text{CC}})U_{\text{C}} - \frac{1}{2}n_{\text{CH}}U_{\text{H}_2} + 4RT - \frac{7}{4}n_{\text{CH}}RT$$

$$\Delta H_{f,T}^{\circ} = V_{\text{steric}} - n_{\text{CH}}(b_{\text{CH}} + \frac{1}{4}U_{\text{C}} + \frac{1}{2}U_{\text{H}_2} + \frac{7}{4}RT) - n_{\text{CC}}(b_{\text{CC}} + \frac{1}{2}U_{\text{C}}) + 4RT$$

Defining $a_{\text{CH}} \equiv -(b_{\text{CH}} + \frac{1}{4}U_{\text{C}} + \frac{1}{2}U_{\text{H}_2} + \frac{7}{4}RT)$ and $a_{\text{CC}} \equiv -(b_{\text{CC}} + \frac{1}{2}U_{\text{C}})$, we have

$$\Delta H_{f,T}^{\circ} = V_{\text{steric}} + n_{\text{CH}}a_{\text{CH}} + n_{\text{CC}}a_{\text{CC}} + 4RT \quad (17.90)$$

One uses (17.90) to find a_{CH} and a_{CC} by a least-squares fit to 25°C experimental ΔH_f° data for several gas-phase hydrocarbons. (Typically, $a_{\text{CH}} \approx -4\frac{1}{2}$ kcal/mol and $a_{\text{CC}} \approx 2\frac{1}{2}$ kcal/mol.) Then Eq. (17.90) can be used to find the gas-phase $\Delta H_{f,298}^{\circ}$ for any saturated hydrocarbon from its steric energy calculated by molecular mechanics. A similar derivation gives an analogous equation for other kinds of compounds. In arriving at (17.90), the contributions of vibrational energy were ignored. It is assumed that these contributions are allowed for when a_{CC} and a_{CH} are fit to 25°C ΔH_f° data. If more than one conformation is significantly populated at 25°C, one uses $\Delta H_f^{\circ} = \sum_i x_i \Delta H_{f,i}^{\circ}$, where x_i is the mole fraction of conformation i as calculated using the enthalpy and entropy differences between conformations, and where $\Delta H_{f,i}^{\circ}$ is calculated from (17.90) for each conformation. In practice, the accuracy of (17.90) is improved by including several correction terms; see Prob. 17.45.

Performance

How well does molecular mechanics perform? When a well-parametrized force field is applied to compounds similar to those used in the parametrization, one can get extremely good results. For monofunctional organic compounds, MM3 typically gives gas-phase ΔH_f° errors that run 0 to 1 kcal/mol (which is the same magnitude as the experimental errors in such data), gives bond lengths within 0.01 Å and bond angles within 2° of experimental values, and usually correctly predicts the most stable conformer.

However, one cannot expect such good performance when molecular mechanics is applied to broad classes of compounds. For a force field with 60 atom types, the number of possible types of IJKL dihedral angles is roughly $\frac{1}{2}(60)(45)(45)(60) \approx 3 \times 10^6$. The factor 45 is used instead of 60 because some atom types (for example, halogens, all H types, carbonyl oxygen) can only occur as end atoms of a molecular dihedral angle. For each of these 3 million types of dihedral angle, one needs V_1 , V_2 , and V_3 values. No available force field will have parameters for all the dihedral angles that might be encountered, so one is frequently faced with the problem of missing parameters. Most molecular-mechanics programs will use some procedure to estimate a value for each missing parameter. If torsion parameters for the IJKL dihedral angle are missing, the program might use torsion parameters for a dihedral angle of type IJKM, with one atom type replaced by a similar atom type (for example, hydrogen bonded to O might be replaced with hydrogen bonded to C). If this fails to produce the torsion parameters, the program might use a value that is typical for torsion about the JK bond, ignoring what types of atoms are bonded to J and K. When such estimated parameters are used in a molecular-mechanics calculation, the reliability of the result may be greatly diminished.

For dealing with noncovalent interactions such as hydrogen bonding and dispersion, the OPLS-AA and the MMFF94s force fields performed better than the MM2, MM3, and AMBER fields [R. S. Paton and J. M. Goodman, *J. Chem. Inf. Model.*, **49**, 944 (2009)]. [MMFF94s differs from MMFF94 only in the treatment of conjugated N atoms such as in amides; T. A. Halgren, *J. Comput. Chem.*, **20**, 720 (1999).]

Programs

The Spartan program (Section 15.14) can do MM energy minimizations with the MMFF and Tripos force fields and does conformational searching. HyperChem (Section 15.14) has in it the force fields MM+ (an extension of MM2), AMBER, BIO+ (a version of CHARMM), and OPLS and can do energy minimization, molecular dynamics, and Monte Carlo calculations. PCModel (www.serenasoft.com) is a Windows and Macintosh program with the force fields MM3, MMFF94, Amber, and OPLS-AA. MacroModel [F. Mohamadi et al., *J. Comput. Chem.*, **11**, 440 (1990); www.schrodinger.com] has the force fields MM2, MM3, AMBER, OPLS-AA, AMBER94, and MMFF. TINKER is a free molecular-mechanics and molecular-dynamics program for Windows, Macintosh, and Linux with the AMBER, CHARMM, MMFF, MM2, MM3, and other force fields (dasher.wustl.edu/tinker/) and has a free graphical interface called Force Field Explorer. The WebMO Demo Server (www.webmo.net/demo) allows you to access Tinker and run MM2, MM3, and OPLS/AA calculations. CHARMM (www.charmm.org) is a workstation program with the CHARMM and MMFF94 force fields. ChemBio3D Ultra (www.cambridgesoft.com) is a Windows molecular-modeling program with MM2. BALLView (www.ballview.org) is a free molecular modeling Windows and Macintosh program with the AMBER and CHARMM force fields [A. Moll et al., *J. Comput.-Aided Mol. Des.*, **19**, 791 (2005); *Bioinformatics*, **22**, 365 (2006)]. *Gaussian 09* can do MM calculations with the AMBER, DREIDING, and UFF force fields. For other molecular-mechanics programs, see en.wikipedia.org/wiki/List_of_software_for_molecular_mechanics_modeling.

QM/MM Methods

Molecular mechanics can treat very large molecules but is not well suited to treat chemical reactions. Quantum-mechanical methods (ab initio, density functional, and semiempirical) can treat chemical reactions but are not well suited to treat very large molecules. To deal with chemical reactions in very large systems, one can use

a combined quantum-mechanical/molecular-mechanics (QM/MM) approach, using quantum mechanics to treat the part of the system most affected by the reaction and molecular mechanics to treat the rest. For a reaction in solution, one treats the reacting solute molecules using QM and the surrounding solvent molecules using MM. For an enzyme-catalyzed reaction, one treats the enzyme's active site and part or all of the substrate molecule(s) using QM and the rest of the enzyme and surrounding water molecules using MM. For enzyme-catalyzed reactions, one must deal with covalent bonds that join the QM and MM regions. Several methods have been used to handle this problem. Any quantum-mechanical method can be used for the QM region. Because of their speed, semiempirical methods such as AM1, PM3, PM6, and SCC-DFTB are more often used, but a growing number of studies use DFT-D.

The effective Hamiltonian for the QM/MM system is written as

$$\hat{H}_{\text{eff}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{QM/MM}} + \hat{H}_{\text{MM}} \quad (17.91)$$

where $\hat{H}_{\text{QM}}$ is the Hamiltonian of the quantum-mechanical region in a vacuum, $\hat{H}_{\text{QM/MM}}$ is the Hamiltonian for the interaction between the QM and MM regions, and $\hat{H}_{\text{MM}}$ is taken as equal to the steric energy V_{MM} [Eq. (17.83)] of the MM region. If ψ_{QM} is the wave function of the QM region, the energy E of the system is taken as

$$E = \langle \psi_{\text{QM}} | \hat{H}_{\text{QM}} + \hat{H}_{\text{QM/MM}} | \psi_{\text{QM}} \rangle + V_{\text{MM}}$$

$\hat{H}_{\text{QM/MM}}$ is often approximated as the electrostatic interactions between the charges (electrons and nuclei) of the QM region and the partial atomic charges on the atoms in the MM region plus the van der Waals interactions between the atoms of the QM region and the atoms of the MM region:

$$\hat{H}_{\text{QM/MM}} = - \sum_i \sum_M \frac{Q_M}{r_{i,M}} + \sum_\alpha \sum_M \frac{Z_\alpha Q_M}{R_{\alpha M}} + \sum_\alpha \sum_M \epsilon_{\alpha M} \left[\left(\frac{R_{\alpha M}^*}{R_{\alpha M}} \right)^{12} - 2 \left(\frac{R_{\alpha M}^*}{R_{\alpha M}} \right)^6 \right]$$

where i and α denote, respectively, the electrons and the nuclei in the QM region and M denotes an atom with partial atomic charge Q_M in the MM region. This approximation neglects polarization of solvent molecules by charges in the QM region. For a fixed configuration of nuclei (Born–Oppenheimer approximation), all terms in $\hat{H}_{\text{QM/MM}}$ are constants except the first double sum. In solving the electronic Schrödinger equation using the Hamiltonian $\hat{H}_{\text{eff}}$, the first double sum in $\hat{H}_{\text{QM/MM}}$ has the effect of extra nuclei with charges Q_M . This produces extra one-electron integrals in the Fock matrix. ψ_{QM} is a function of the electronic coordinates of the QM region, and both ψ_{QM} and E are parametric functions of the nuclear coordinates of the QM region and the coordinates of the atoms in the MM region. Taking minus the partial derivatives of E with respect to the nuclear coordinates, we get the forces needed for a molecular-dynamics simulation.

The *pseudobond* QM/MM method [Y. Zhang, T.-S. Lee, and W. Yang, *J. Chem. Phys.*, **110**, 46 (1999)] divides a large molecule into two parts by cutting a covalent bond X—Y between atoms X and Y and does the QM part of the calculation using an ab initio method. Let atom X be in the part of the molecule to be treated by QM, and suppose that Y is a carbon atom. The free valence created by cutting the bond is “capped” by singly bonding atom X to a one-free-valence pseudocarbon atom C_{ps} that has seven valence electrons (so as to be one electron short of an octet), a nuclear charge of seven, and no core electrons. The effect of the core electrons is handled by using an effective core potential (Section 15.16). The pseudobond X— C_{ps} mimics the properties of the X—Y bond. Tests with the QM calculation done with the HF and DFT methods gave good results.

For reviews of the QM/MM method, see J. Gao in K. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry*, vol. 7, VCH, 1996, Chapter 3; H. Lin and D. G. Truhlar, *Theor. Chem. Acc.*, **117**, 185 (2007).

17.6 Empirical and Semiempirical Treatments of Solvent Effects

Explicit-Solvent versus Continuum-Solvent Methods

Theoretical treatments of solvation can be categorized as either explicit-solvent approaches, in which many individual solvent molecules are explicitly included, or as continuum-solvent methods, where the solvent molecules are replaced by a continuous dielectric (Section 15.17).

In an explicit-solvent treatment, one applies the molecular dynamics or the Metropolis Monte Carlo method (Sections 15.11 and 17.5) to a system of a solute molecule (or molecules if a chemical reaction is being studied) surrounded by hundreds or thousands of solvent molecules; by suitable averaging, one obtains thermodynamic or kinetic properties. The solvent molecules are treated using an empirical force field. Some of the models used to represent water molecules are discussed in *Leach*, Section 4.14. The solute molecule(s) can be modeled using molecular mechanics or semiempirical quantum mechanics (the QM/MM method of Section 17.5).

Continuum-solvent methods can be categorized as either classical or quantum-mechanical.

Classical Continuum-Solvent Methods

The simplest and crudest continuum-solvent method is the **solvent-accessible-surface area** (SASA) model, which assumes that the standard free-energy of solvation $\Delta G_{\text{solv}}^\circ$ can be expressed as

$$\Delta G_{\text{solv}}^\circ = \sum_i \sigma_i A_i \quad (17.92)$$

Recall (Section 15.7) that the van der Waals surface of a molecule is the outer surface formed by intersecting atomic spheres with van der Waals radii. The **solvent-accessible surface** is the surface traced out by the center of a spherical solute molecule as it rolls on the molecular van der Waals surface. For H_2O , a sphere of radius 1.4 Å is traditionally used in (17.92). A_i is the **solvent-accessible surface area** of atom i or group i , depending on whether an atom-based or group-based approach is used. The sum goes over all atoms (or groups) in the molecule. A_i is that portion of the surface area of a sphere centered at atom i and having radius $r_i + r_{\text{solvent}}$ (where these quantities are the van der Waals radius of atom i and the solvent molecular radius) that is not contained within any of the corresponding spheres centered at the other atoms in the molecule. The σ_i 's are empirical parameters called **atomic surface tensions** or **atomic solvation parameters**. They have the same dimensions as surface tension but are not macroscopic surface tensions. The σ_i 's are found by fitting known $\Delta G_{\text{solv}}^\circ$ 25°C data for the solvent under study. The SASA method is especially useful for rapid estimation of $\Delta G_{\text{solv}}^\circ$ of biopolymers.

A related but more elaborate method than the SASA model is the SM5.0R method [G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. B*, **101**, 7147 (1997); Hawkins et al., *J. Org. Chem.* **63**, 4305 (1998)]. Here, (17.92) is used, but A_i is taken as the exposed van der Waals surface area of atom i (rather than the SASA of i) and σ_i , instead of being a constant for a given atom in a given solvent, has the form $\sigma_i = \sigma_i^{(0)} + \sum_{j \neq i} \sigma_{ij} f(R_{ij})$.

Here, the sum goes over all atoms in the solute molecule except atom i , $\sigma_i^{(0)}$ and σ_{ij} are parameters whose values depend on the nature of atoms i and j , and $f(R_{ij})$ is a certain function of the distance R_{ij} between atoms i and j . The function f contains parameters and is defined to be zero for all distances greater than a certain cutoff distance. Inclusion of the sum allows σ_i to be affected by the environment of atom i in the molecule. The method contains about 40 parameters for water as the solvent and has been parametrized for compounds of H, C, N, O, S, F, Cl, Br, and I in water and in organic solvents.

Rather than develop a separate set of parameter values for each organic solvent, SM5.0R takes each surface tension coefficient $\sigma_i^{(0)}$ and σ_{ij} to depend on certain properties of the solvent such as its index of refraction, surface tension, and fraction of nonhydrogen atoms in the solvent molecule that are aromatic carbons, and to depend on parameters many of whose values depend on the nature of atom i or atoms i and j (whether they are H, C, N . . .) but whose values are the same for every organic solvent. The values of the parameters are fitted using a training set of 1836 $\Delta G_{\text{solv}}^\circ$ values for 227 uncharged solutes in 90 organic solvents.

The molecular geometry used to calculate R_{ij} is that of the gas-phase solute molecule. Test calculations on small uncharged solutes in water using AM1 geometries gave excellent results, but for charged solutes the performance was not so good. The SM5.0R method is useful for rapid estimations for large molecules. [The SM5 stands for solvation model version 5; the R indicates that rigid (that is, gas-phase rather than solution-phase) geometries are used; the 0 indicates that no use is made of atomic charges.]

Another classical method to estimate $\Delta G_{\text{solv}}^\circ$ is the **generalized Born/surface area** (GB/SA) method of Still and co-workers [W. C. Still et al., *J. Am. Chem. Soc.*, **112**, 6127 (1990); D. Qiu et al., *J. Phys. Chem. A*, **101**, 3005 (1997)]. In 1920, Born showed that when a charge Q distributed on the surface of a conducting sphere of radius a is transferred from vacuum to a dielectric medium with dielectric constant ϵ_r , the free-energy change is $\Delta G = -\frac{1}{2}(1 - 1/\epsilon_r)Q^2/4\pi\epsilon_0 a$. Now imagine a system of several charged spheres having charges Q_i and radii a_i , with R_{ij} being the distance between spheres i and j . If such a system is transferred from vacuum to a medium of dielectric constant ϵ_r , the free-energy change can be approximated by adding the Born expression $-\sum_i \frac{1}{2}(1 - 1/\epsilon_r)Q_i^2/4\pi\epsilon_0 a_i$ to the interchange electrostatic potential-energy change $\sum_{j>i} \sum_i Q_i Q_j / \epsilon_r 4\pi\epsilon_0 R_{ij} - \sum_{j>i} \sum_i Q_i Q_j / 4\pi\epsilon_0 R_{ij}$ to get

$$\Delta G \approx -\left(1 - \frac{1}{\epsilon_r}\right) \sum_i \sum_j \frac{Q_i Q_j}{4\pi\epsilon_0 R_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\epsilon_r}\right) \sum_i \frac{Q_i^2}{4\pi\epsilon_0 a_i} \quad (17.93)$$

an equation sometimes called the *generalized Born equation*. If we model a solute molecule as a set of overlapping spheres with charges Q_i (where Q_i is a partial atomic charge) and radii a_i , then the generalized Born equation gives an approximation to the electrostatic contribution $\Delta G_{\text{solv,el}}^\circ$ (Section 15.17) to the solvation free energy. However, this method would overestimate the interactions between the solvent and charges buried within the solute molecule's interior. Instead, Still and co-workers proposed replacing (17.93) with

$$\Delta G_{\text{solv,el}}^\circ = -\frac{1}{2} \left(1 - \frac{1}{\epsilon_r}\right) \sum_i \sum_j \frac{Q_i Q_j}{4\pi\epsilon_0 [R_{ij}^2 + a_i a_j e^{-R_{ij}/4a_i a_j}]^{1/2}} \quad (17.94)$$

where the sums go over all atoms in the molecule. For two charges, this expression reduces to the Born equation when $R_{ij} = 0$, gives a result close to the Onsager result for a dipole in a sphere (Section 15.17) when $R_{ij} \ll a_i a_j$, and gives a result close to the sum of the Born plus interchange potential energy when $R_{ij} \gg a_i a_j$. The partial atomic charges Q_i to be used in (17.94) can be found by any of the usual methods used in molecular mechanics,

but the best results are obtained using either charges fit to molecular electrostatic potentials or OPLS charges, which were derived from liquid simulations. The Born atomic radii a_i are calculated by a complicated procedure [see *Leach*, Chapter 11; D. Qiu et al., *J. Phys. Chem. A*, **101**, 3005 (1997)]. The term *Born radius* for a_i is misleading, in that a_i in (17.94) is some sort of average distance from the atomic charge Q_i to the dielectric continuum that surrounds the solute molecule.

Still and co-workers modeled the cavitation and van der Waals contributions to $\Delta G_{\text{solv}}^\circ$ by

$$\Delta G_{\text{cav}}^\circ + \Delta G_{\text{vdw}}^\circ = \sum_k \sigma_k A_k \quad (17.95)$$

where the sum goes over all atoms in the solute molecule, A_k is the solvent-accessible surface area for atom k , and the parameter σ_k is taken as 10 cal/(mol Å²) for S and sp^3 C atoms, 7 cal/(mol Å²) for P and sp^2 and sp C atoms, and 0 for O and N atoms. In (17.95), a united-atom approach (Section 17.5) is used for hydrogens, so there is no σ_k for H.

Addition of (17.94) and (17.95) gives the GB/SA expression for $\Delta G_{\text{solv}}^\circ$. The GB/SA expression is readily differentiated analytically, making it easy to use in molecular-mechanics energy minimizations (one finds the geometry that minimizes the sum of the MM steric energy and $\Delta G_{\text{solv}}^\circ$), and molecular-dynamics and Metropolis Monte Carlo simulations with inclusion of solvent effects. The GB/SA method is available in the MacroModel program (Section 17.5).

Quantum-Mechanical Continuum Solvation Models

Several ab initio continuum solvation models were discussed in Section 15.17. One can calculate $\Delta G_{\text{solv,el}}^\circ$ by such SCRF methods as the dipole-in-a-sphere, the multipole expansion, or the PCM methods using semiempirical methods such as AM1 or PM3 instead of an ab initio electronic-structure method. Thus the programs MOPAC2012 and AMPAC 9 implement the COSMO method.

Cramer, Truhlar, and co-workers developed a series of semiempirical continuum solvation models SM x , where SM stands for solvation model and $x = 1, 2, 3, 4, 5, 6, 8$, D is the version. The models SM1, SM2, SM3, and SM4 are obsolete. The SM x methods are quantum-mechanical versions of Still's GB/SA method, except for SMD. The earlier versions of SM5 (SM5.2 and SM5.4) were parametrized for use with the AM1 and PM3 semiempirical methods, but later versions of SM5 (SM5.42 and SM5.43) were also parametrized for use with the DFT and HF methods; SM6 and SM8 [A. V. Marenich et al., *J. Chem. Theory Comput.*, **3**, 2011 (2007)] are parametrized for use only with DFT or HF calculations. In SM x , the solvation energy is written as

$$\Delta G_{\text{solv}}^\circ = \Delta E_{\text{EN}} + G_{\text{p}} + \Delta G_{\text{CDS}}^\circ \quad (17.96)$$

In this equation, the electronic and nuclear term ΔE_{EN} is the change in the electronic energy including nuclear repulsion of a solute molecule when it goes from the gas phase into solution; ΔE_{EN} corresponds to $\langle \psi^{(f)} | \hat{H}^0 | \psi^{(f)} \rangle - \langle \psi^{(0)} | \hat{H}^0 | \psi^{(0)} \rangle$ in Eq. (15.84). The polarization contribution G_{p} is the free-energy change due to electrostatic interactions between solute and solvent, including polarization of the solvent. G_{p} corresponds to $\langle \psi^{(f)} | V^{\text{int}} | \psi^{(f)} \rangle + E_{\text{pol}}$ in Section 15.17. The sum $\Delta E_{\text{EN}} + G_{\text{p}}$ equals $\Delta G_{\text{solv,el}}^\circ$ of (15.84). The cavitation–dispersion–solvent-disposition term G_{CDS}° is the free-energy change due to creation of the cavity around the solute molecule, dispersion interactions between solute and solvent molecules, and structural changes in the solvent, due to such things as solute–solvent hydrogen bonding. Other contributions are small and are neglected.

The SM5 series of models [G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem.*, **100**, 19824 (1996) and references cited therein] has been parametrized for compounds containing H, C, N, O, F, S, Cl, Br, and I. In SM5, the polarization contribution

G_p is taken as equal to the Still expression (17.94), except that the 4 in the exponential is replaced by 4.2 when the two atoms i and j are C and H. The Born radii a_i are calculated using Still's procedure or a faster procedure called pairwise descreening (PD) that is an approximation to Still's procedure [Hawkins et al., *J. Phys. Chem.*, **100**, 19824 (1996)].

In the version SM5.2PD, the atomic partial charges are taken as the Mulliken-population-analysis net charges (Section 15.6), which are given by $C_B - \sum_r P_{rr}$, where C_B is the charge on the core of atom B, the sum goes over all valence AOs centered on B, and P_{rr} is a density matrix element [Eq. (14.42)]. The SM5 model does an AM1 or PM3 or DFT or ab initio HF SCF calculation to minimize the sum of the molecular electronic energy, including nuclear repulsion and G_p . Because G_p is included in the quantity to be minimized, one must modify the Fock matrix elements (14.41) by adding a term that is closely related to G_p and that contains the atomic partial charges Q_i that occur in G_p in (17.94). To do the SCF calculation, one starts with a guess for the density matrix elements P_{rs} , uses these to calculate the initial estimates of the atomic charges Q_i , and uses the initial Q_i 's and P_{rs} 's to calculate the initial estimates of the Fock matrix elements. Solution of the Fock–Roothaan equations then allows improved density matrix elements to be found, which are used to start another cycle of the calculation. The calculation ends when self-consistency is reached.

The versions SM5.4 and SM5.4PD use improved charges (called class IV charges) instead of the Mulliken charges (called class II charges) used in SM5.2PD. Class IV charges are found as follows. For a training set of molecules with known dipole moments, one calculates the Mulliken charges Q_i using the AM1 or PM3 method and writes $Q_{i,IV} = g(Q_i)$, where g is a certain function that contains parameters and that maps the Mulliken charges into class-IV charges. One varies the parameters in g so that the molecular dipole moments calculated from the class-IV charges give a good fit to the known dipole moments. Once the parameters in g have been determined, one can use g to calculate class-IV charges from the Mulliken charges for any molecule. This particular method of getting class-IV charges is called CM1 (charge model 1).

The version SM5.42R [T. Zhu et al., *J. Chem. Phys.*, **109**, 9117 (1998)] uses the improved method CM2 of generating the class-IV charges (hence the 2 in 5.42). The version SM5.43R [J. D. Thompson et al., *J. Phys. Chem. A*, **108**, 6532 (2004)] uses the further-improved CM3 method to generate the class-IV charges.

The version SM6 [C. P. Kelly et al., *J. Chem. Theory Comput.* **1**, 1133 (2005)] improves on SM5.43R by using charges that are based on the CM4 method and by using a larger and more varied data set for the parametrization.

The SM5 and SM6 models calculate G_{CDS}° using the SASA approach:

$$G_{CDS}^\circ = \sum_k \sigma_k A_k$$

where A_k is the solvent-accessible surface area of atom k , and σ_k has a form similar to that of σ_i in the SM5.0R model discussed near the beginning of this subsection.

The SMD model (D stands for density), unlike the preceding models, does not use partial atomic charges but instead is based on the interaction of the electron charge density $-\rho$ of the solute molecule, with the solvent represented as a continuum [A. Marenich et al., *J. Phys. Chem. B*, **113**, 6378 (2009)]. The method can be used for any solute and for any solvent for which the dielectric constant, refractive index, surface tension, and acidity and basicity parameters are known, and can be applied with any quantum-mechanical method to calculate ρ . The method was parametrized using data for many solutes in several solvents and using DFT and HF calculations. The method gives good results with mean absolute errors in 25°C ΔG_{solv}° values of 0.6 to 0.9 kcal/mol for a variety of solutes and solvents, and it outperforms the default solvation models in several standard quantum-chemistry programs.

Various SMX solvation models are available in many quantum-chemistry programs. For listings, see comp.chem.umn.edu/solvation.

COSMO-RS (COSMO for real solvents) is an extension of COSMO (Section 15.17) beyond the dielectric-continuum approximation [A. Klamt et al., *J. Phys. Chem. A*, **102**, 5074 (1998)]. COSMO-RS does not treat the solvent as a continuous dielectric, but treats solute and solvent molecules on an equal footing and uses cavities and surface charges for both the solute and solvent molecules. COSMO-RS has eight general parameters plus two parameters for each different element and does not use solvent-specific parameters. Thus for solutions containing compounds of H, C, N, O, and Cl, COSMO-RS has 18 parameters, whose values were chosen by fitting to experimental data.

17.7 Chemical Reactions

Because of the difficulties involved in ab initio calculations of potential-energy surfaces for reactions, it would be highly desirable to have a semiempirical method that gives reliable reaction PES results. The MNDO, AM1, and PM3 methods have been widely applied to calculate relevant portions of PESs for chemical reactions (often with inclusion of solvent effects) so as to elucidate reaction mechanisms and transition-state structures.

An analysis of MNDO results for 24 simple organic reactions found that MNDO usually gave a fairly realistic transition-state structure [S. Schröder and W. Thiel, *J. Am. Chem. Soc.*, **107**, 4422 (1985)], with average absolute deviations from ab initio SCF MO results of 0.06 Å in bond lengths, 8° in bond angles, and 11½° in dihedral angles. However, the calculated MNDO barrier heights were not accurate, with an average absolute deviation of 22 kcal/mol from ab initio MP4 or MP3 results. MNDO was suggested as being suitable “for fast initial scans of potential surfaces to assess their qualitative features, but it has to be kept in mind that some of these characteristics may change at higher levels” (Schröder and Thiel, op. cit.).

Some shortcomings of AM1 for studying reactions are discussed by O. N. Ventura in S. Fraga (ed.), *Computational Chemistry*, Part B, Elsevier, 1992, pp. 605–607. Ventura concludes that “there is no semiempirical method at present which can be used reliably in all situations [to study reactions].”

A method that improves the accuracy of a semiempirical PES for a particular reaction is NDDO-SRP of Truhlar and co-workers (NDDO with specific reaction parameters; recall that MNDO, AM1, and PM3 are all NDDO methods). Here, some of the parameters in the AM1 method are adjusted either to reproduce such experimental data as the reaction’s energy change, barrier height, and rate constant or to reproduce a small number of points on the PES calculated with an ab initio method that includes correlation. See the references cited in W. Thiel, *Adv. Chem. Phys.*, **93**, 731 (1996).

Although the force fields mentioned in Section 17.5 do not apply to bond-breaking reactions, they can be applied to reactions involving only a change in molecular conformation. Some attempts have been made to develop force fields that will describe bond breaking; see F. Jensen and P.-O. Norrby, *Theor. Chem. Acc.*, **109**, 1 (2003); Cramer, Section 2.4.2.

MO theory has been used to draw qualitative conclusions about the course of chemical reactions. The most fruitful applications have come from the **Woodward–Hoffmann rules**, which predict the preferred path and stereochemistry for many important classes of organic reactions. As an example of the application of these rules, we consider the cyclization of a substituted *s-cis*-butadiene to a substituted cyclobutene. The reaction can take two possible steric courses described as *conrotatory* or *disrotatory*, depending on whether the terminal groups rotate in the same or opposite senses as the reaction proceeds. Note the difference in products in [Fig. 17.11](#).

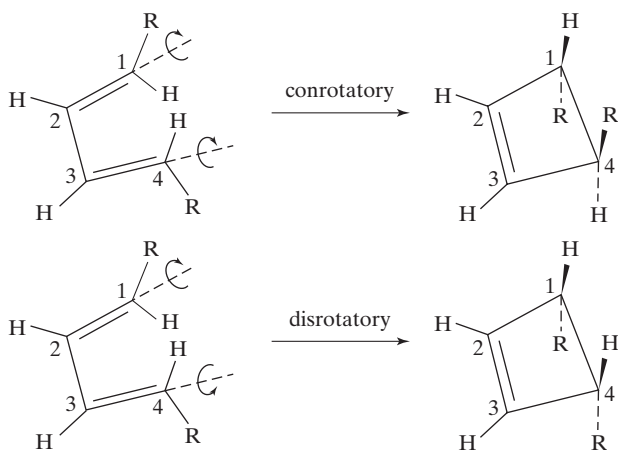


FIGURE 17.11 Conrotatory and disrotatory cyclizations. All atoms lie in the same plane except those shown with dashed or heavy bonds.

When the reaction is thermally induced, the butadiene cyclization process is found to be conrotatory, but when the reaction is photochemically induced, the process is disrotatory. The simplest approach that explains these facts starts with the assumption that the energy change during the reaction is determined primarily by the energy change in the highest-occupied MO (HOMO) of the substituted *s-cis*-butadiene molecule. This is the π MO ϕ_2 in Fig. 17.1. **Figure 17.12** shows that for ϕ_2 , a conrotatory motion of the π AOs on carbons 1 and 4 causes the positive lobe of $C_1 2p\pi$ to overlap the positive lobe of $C_4 2p\pi$; this overlap gives a bonding interaction of these AOs and leads to formation of the 1–4 σ bond of the cyclobutene. On the other hand, a disrotatory motion of these two AOs causes overlap of the positive lobe of one π AO with the negative lobe of the other π AO; this gives an antibonding interaction. Thus we expect the conrotatory process to have a lower activation energy than the disrotatory and to be preferred over the disrotatory. As a further check, we note that conrotatory motion leads to an antibonding interaction between the carbon 1 and 4 π AOs in the lowest π MO ϕ_1 ; this leaves the π AOs of carbons 2 and 3 in ϕ_1 to form the 2–3 π bond in the cyclobutene. (Note that if the butadiene is less symmetrically substituted than in Fig. 17.11, clockwise and counterclockwise conrotatory motions lead to a mixture of two products.)

When the above reaction is induced photochemically, an absorbed photon excites an electron from ϕ_2 to the butadiene π MO ϕ_3 (Fig. 17.1). The highest occupied MO is now ϕ_3 , in which the $2p\pi$ AOs on carbons 1 and 4 have the same phases (rather than opposite phases as in ϕ_2) and in which a disrotatory motion produces positive

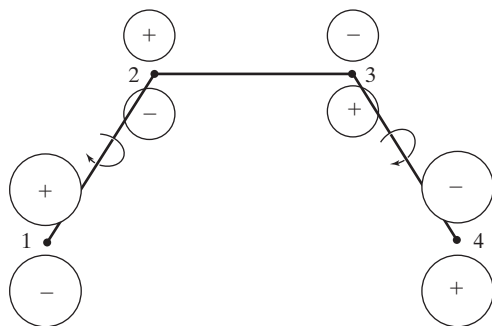


FIGURE 17.12 Relative phases of AOs in the highest occupied butadiene MO.

bonding overlap between these AOs. Thus we predict disrotatory ring closure, as observed. The same reasoning correctly predicts the courses of other polyene ring closures (Prob. 17.46).

Provided the reactants and products do not differ greatly in energy, a high barrier for the forward reaction implies a high barrier for the reverse reaction (in which the reaction path is traversed in the opposite direction). Thus the above reasoning also applies in determining the course of the reverse, ring-opening, reactions.

Rather than considering only the energy change in the HOMO of butadiene, one can use a less approximate approach that looks at the energy changes in all MOs (occupied or unoccupied) that are involved in the bonds being broken or formed and that uses symmetry to correlate the MOs of the reactant(s) with the MOs of the product(s); see *Lowe and Peterson*, Section 14-9.

For further details of the Woodward–Hoffmann rules and their application to a wide variety of organic reactions, see R. B. Woodward and R. Hoffmann, *Angew. Chem. Intern. Ed.*, **8**, 781 (1969); *The Conservation of Orbital Symmetry*, Academic Press, 1970; R. E. Lehr and A. P. Marchand, *Orbital Symmetry*, Academic Press, 1972.

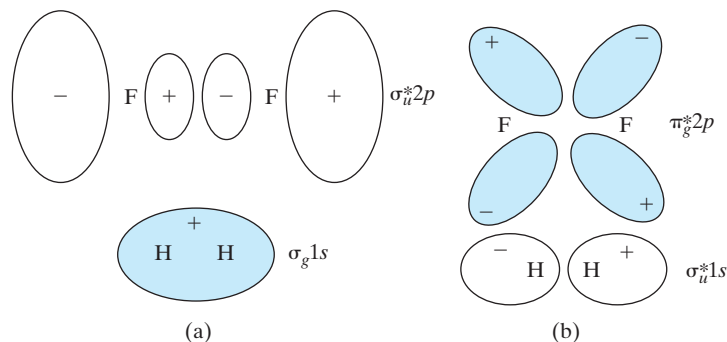
Pearson has applied orbital symmetry concepts to inorganic reactions. As two reactant molecules approach, electrons begin to flow from the HOMO of one to the lowest unoccupied MO (LUMO) of the other. These two MOs are called **frontier orbitals**. For a low activation energy, we require a positive overlap between these two MOs.

An example is the $\text{H}_2 + \text{F}_2 \rightarrow 2\text{HF}$ reaction. Consider a proposed mechanism in which the two molecules collide broadside to give a four-center transition state. The HOMO of H_2 is the $\sigma_g 1s$ MO; the LUMO of F_2 is the $\sigma_u^* 2p$ MO (Sections 13.6 and 13.7). Flow of electrons out of H_2 $\sigma_g 1s$ toward F_2 $\sigma_u^* 2p$ would lead to breaking of the H—H bond and formation of two H—F bonds. However, **Fig. 17.13a** shows that these two MOs do not have a positive overlap. Hence electron flow from H_2 to F_2 is forbidden by symmetry. Figure 17.13b shows the HOMO of F_2 and LUMO of H_2 . Here there is a positive overlap, but flow of electrons out of the antibonding π_g^* MO would strengthen rather than weaken the F—F bond. We conclude that this bimolecular one-step mechanism has a high activation energy and is not favored. (The same reasoning applies to the famous $\text{H}_2 + \text{I}_2$ reaction.)

Further applications of orbital symmetry to inorganic reactions may be found in R. G. Pearson, *Acc. Chem. Res.*, **4**, 152 (1971); *J. Am. Chem. Soc.*, **94**, 8287 (1972); *Symmetry Rules for Chemical Reactions*, Wiley, 1976.

The frontier-orbital approach sometimes fails. For a criticism of the frontier-orbital theory, see M. J. S. Dewar, *THEOCHEM*, **59**, 301 (1989).

FIGURE 17.13 HOMOs and LUMOs for H_2 and F_2 . Occupied MOs are shaded.



17.8 The Future of Quantum Chemistry

In the 1950s there was a general belief that meaningful *ab initio* calculation of molecular properties for all except very small molecules was out of the question. Quantum-chemistry books written in this period contain such statements as “we cannot hope ever to make satisfactory *ab initio* calculations [for organic compounds]” and “it is wise to renounce at the outset any attempt at obtaining precise solutions of the Schrödinger equation for systems more complicated than the hydrogen molecule ion.” In 1959, Mulliken and Roothaan identified the “bottleneck” holding up accurate quantum-mechanical calculations on polyatomic molecules as the difficulty in evaluating multicenter integrals. This bottleneck has now been eliminated.

Ab initio HF calculations and geometry optimizations on moderate-size molecules have become routine, and computationally efficient methods (for example, DFT, MP2, and SCS-MP2) for inclusion of electron correlation are available. The degree of reliability of various quantum-mechanical methods and basis sets has been established by numerous calculations. The size of a molecule for which one can do an accurate *ab initio* or density-functional calculation is limited by the speed and storage capacity of the available electronic computers. As larger and faster computers are developed, it will become feasible to treat larger molecules.

The very substantial progress in quantum chemistry in recent years has made quantum-mechanical calculations a valuable tool to help decide a wide variety of questions of real chemical interest. Whereas years ago quantum-mechanical calculations on molecules were largely confined to journals read mainly by theoretical chemists, nowadays such calculations appear routinely in journals such as the *Journal of the American Chemical Society*, read by all kinds of chemists. Quantum chemistry is being applied to such problems as the hydration of ions in solution, surface catalysis, the structures and energies of reaction intermediates, the conformations of biological molecules, and the study of enzyme-catalyzed reactions. In many cases, theoretical calculations may not give definitive answers, but they are frequently good enough to allow for a very fruitful interaction of theory and experiment. Moreover, qualitative concepts such as the Woodward–Hoffmann rules have provided considerable insight into the course of chemical reactions and into chemical bonding.

The 1998 Nobel Prize in chemistry was shared by Walter Kohn, one of the developers of density-functional theory, and John A. Pople, one of the developers of the *Gaussian* series of programs and widely used Gaussian basis sets, the Pariser–Parr–Pople method, and the CNDO and INDO methods, and one of the first to apply the MP and CC methods to molecular calculations. The Nobel committee noted that computational quantum chemistry is “revolutionising the whole of chemistry.”

In 1929, Dirac wrote, “The underlying physical laws necessary for the mathematical theory of . . . the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble.” Application of high-speed digital computers to quantum chemistry has overcome to a significant degree the difficulties referred to by Dirac. Of course, just a small fraction of chemically important problems have been successfully treated by quantum mechanics, but future prospects are bright.

Ab initio, density-functional, and semiempirical calculations are now routinely used by many chemists as a valuable guide to experimental work and are revolutionizing the way chemistry is done. The future of quantum chemistry and the future of chemistry are inextricably linked.

Problems

Sec.	17.2	17.3	17.4	17.5	17.7
Probs.	17.1–17.24	17.25	17.26–17.36	17.37–17.45	17.46–17.47

- 17.1** For the allyl radical $\cdot\text{CH}_2-\text{CH}=\text{CH}_2$, find (a) the HMOs and energies; (b) the mobile bond orders; (c) the π -electron charges; (d) the free valences (Prob. 17.19); (e) the delocalization energy.
- 17.2** Calculate the quantities (a) through (e) of Prob. 17.1 for the allyl cation and anion, $[\text{CH}_2\text{CHCH}_2]^+$ and $[\text{CH}_2\text{CHCH}_2]^-$. Which ion is predicted to be more stable?
- 17.3** For the polyenes (17.28), observed longest-wavelength electronic absorption bands are at 162.5, 217, 268, 304, 334, 364, 390, 410, and 447 nm for $s = 0, 1, 2, 3, 4, 5, 6, 7$, and 9, respectively. Compare these values with those given by the Hückel equation (17.31) and calculate the average absolute percent error.
- 17.4** Verify the butadiene HMO coefficients for ϕ_2 , ϕ_3 , and ϕ_4 in (17.26).
- 17.5** (a) Verify that the HMO coefficients (17.30) satisfy the HMO set of simultaneous equations. [Hint: Use the identity $\sin a + \sin b = 2 \sin \frac{1}{2}(a + b) \cos \frac{1}{2}(a - b)$.] (b) Verify that the coefficients (17.30) give a normalized HMO. To evaluate the needed sum, express the sine function as exponentials and then use the formula for the sum of a geometric series.
- 17.6** Calculate the HMO total bond orders and the π -electron charges for the lowest excited state of butadiene.
- 17.7** Since only the topology of the carbon framework is of significance in the HMO method, the full symmetry of *s-trans*-butadiene need not be used to get the maximum simplification possible in the HMO method; instead, it is sufficient to use only the C_2 axis. (a) Write down the two possible symmetry species for the group $\mathcal{C}_2$. (b) Construct π -electron symmetry orbitals for butadiene, classifying them according to the symmetry species of $\mathcal{C}_2$. (c) Set up and solve the two Hückel secular equations for butadiene using the symmetry orbitals of (b) as basis functions.
- 17.8** Verify that the geometric construction of Fig. 17.5 gives the correct HMO energies of the cyclic polyene C_nH_n .
- 17.9** (a) Calculate the energy needed to compress three carbon–carbon single bonds and stretch three carbon–carbon double bonds to the benzene bond length 1.397 Å. Assume a harmonic-oscillator potential-energy function for bond stretching and compression. Typical carbon–carbon single- and double-bond lengths are 1.53 and 1.335 Å; typical stretching force constants for carbon–carbon single and double bonds are 500 and 950 N/m. (b) Use the result of (a) to calculate an improved β value for benzene from the data following Eq. (17.51).
- 17.10** Typical bond energies in kcal/mol are 99 for C–H, 83 for C–C, and 146 for C=C. The heat of formation of gas-phase benzene from gas-phase hydrogen and carbon atoms is -1323 kcal/mol. Calculate the “experimental” delocalization energy of benzene using these data, first omitting the strain-energy correction (Prob. 17.9) and then including it. (Because bond energies vary from compound to compound, this procedure is very rough.)
- 17.11** (a) Calculate the quantities (a) through (e) of Prob. 17.1 for the diradical trimethylenemethane $\text{C}(\text{CH}_2)_3$. (The degenerate MOs can be taken as either real or complex, similar to the benzene degenerate MOs.) (b) Do the same as in (a) for the propargyl diradical $\text{HC}=\text{C}=\text{CH}$. Save time by using the results of Prob. 17.1. Note that this linear molecule has two sets of spatially perpendicular π MOs.
- 17.12** The ionization energies of the first few polyacenes are 9.4 eV for benzene, 8.3 eV for naphthalene, 7.6 eV for anthracene, and 7.0 eV for tetracene. Use these data to calculate values of α and β in the HMO method. Compare the result with the β value in (17.49). Predict the ionization energy of pentacene. The HMO x values for the highest-occupied MOs of these polyacenes are -1.00 , -0.618 , -0.414 , -0.295 , and -0.220 .
- 17.13** The terms involving α always cancel in calculations of the delocalization energy, so Hess and Schaad measured energies relative to α . In this problem, α is therefore omitted.

The Hess–Schaad $E_{\pi,b}$ values for the various conjugated-bond types are 2.0000β for $\text{CH}_2=\text{CH}$, 2.0699β for $\text{CH}=\text{CH}$, 2.0000β for $\text{CH}_2=\text{C}$, 2.1083β for $\text{CH}=\text{C}$, 2.1716β for $\text{C}=\text{C}$, 0.4660β for $\text{CH}-\text{CH}$, 0.4362β for $\text{CH}-\text{C}$, 0.4358β for $\text{C}-\text{C}$. (a) Verify that for 1,3-butadiene, $\sum_b n_b E_{\pi,b} = 4.466\beta$, which is very close to the Hückel E_π value 4.472β in (17.50). The Hess–Schaad method makes the delocalization energies of noncyclic conjugated polyenes essentially zero. (b) Verify that for benzene, $\sum_b n_b E_{\pi,b} = 7.6077\beta$ and that the Hess–Schaad REPE for benzene is $0.065|\beta|$. (c) Find the Hess–Schaad resonance energy and REPE for each of the following and predict whether the compound will be aromatic, nonaromatic, or antiaromatic: cyclobutadiene; planar [8]annulene; planar [18]annulene; azulene, for which the x values in (17.20) of the occupied MOs are $-2.3103, -1.6516, -1.3557, -0.8870$, and -0.4773 .

- 17.14** Estimate the carbon–carbon bond length in (a) C_5H_5^- ; (b) C_7H_7^+ ; (c) $\text{C}_8\text{H}_8^{2-}$.
- 17.15** (a) Verify the secular equation given in the text for the b_{3u} naphthalene MOs. (b) Set up and solve the a_u, b_{2g} , and b_{1g} HMO secular equations for naphthalene. (c) Find the coefficients of the lowest naphthalene HMO.
- 17.16** Derive Eq. (17.58) for E_π . *Hint:* Start with E_π as the sum of orbital energies and use $e_i = \langle \phi_i | \hat{H}^{\text{eff}} | \phi_i \rangle$.
- 17.17** (a) For p_{rs}^{tot} equal to 1 and to 3, compare the bond-length predictions of Eq. (17.57) with experimental carbon–carbon single- and triple-bond lengths. (b) Look up or use a program to calculate the Hückel bond orders of azulene and compare the predicted bond lengths with the experimental values. [The experimental data can be found in R. J. Buenker and S. D. Peyerimhoff, *Chem. Phys. Lett.*, **3**, 37 (1969).]
- 17.18** Verify that the E_π expression (17.58) holds for 1,3-butadiene.
- 17.19** The *free-valence index* F_r for carbon atom r in a planar conjugated compound is defined as $F_r \equiv \sqrt{3} - \sum'_s p_{rs}$, where the sum is over the atoms bonded to atom r . [The quantity $\sqrt{3}$ is the value of $\sum'_s p_{rs}$ for the central atom of the diradical $\text{C}(\text{CH}_2)_3$, whose central C has the largest possible value for this sum of any trigonally bonded carbon.] F_r is a measure of the unused bonding power of atom r and is used to estimate the susceptibility of a given atom to attack by an uncharged free radical. Calculate F_r for each carbon in 1,3-butadiene, and state which carbon is predicted to be preferentially attacked by free radicals.
- 17.20** For each of the following molecules, put stars on some of the carbon atoms such that starred carbons are bonded only to unstarred carbons and unstarred carbons are bonded only to starred carbons. It may not be possible to do this in all cases. (a) benzene; (b) naphthalene; (c) azulene.
- 17.21** (a) Show that the equations satisfied by the HMO coefficients of a conjugated hydrocarbon are $x_i c_{ri} + \sum_{s \rightarrow r} c_{si} = 0$, $r = 1, \dots, n_C$, where the sum is over carbons bonded to carbon r . If the hydrocarbon is an alternant, then we can divide the carbons into two sets such that carbons in one set are bonded only to carbons in the other set. Verify that for an alternant hydrocarbon, if we replace x_i by $-x_i$ and multiply the coefficients of one set of carbons by -1 in each HMO equation of (a), we obtain equations that are satisfied. Hence (provided $x_i \neq 0$), for each HMO of an alternant hydrocarbon with $(\alpha - e_i)/\beta = x_i$, there is an HMO with $(\alpha - e_i)/\beta = -x_i$, whose coefficients are obtained by multiplying the coefficients of one set of carbons in the first HMO by -1 .
- 17.22** (a) What is the HMO delocalization energy of cyclobutadiene? (b) Calculate the HMO total bond orders in cyclobutadiene.
- 17.23** Show that $\sum_r q_r = n_\pi$ in the HMO method.
- 17.24** (a) Show that, with inclusion of overlap between carbons bonded to each other (the Wheland method), the HMO equation for benzene is like (17.32) except that each x is replaced by w , where $w \equiv (\alpha - e_i)/(\beta - S e_i)$. Hence the values of w are the same as those found for x with overlap omitted. Provided a common value of S is used for any two bonded carbons in a molecule, the same situation holds for any planar conjugated hydrocarbon. (b) Verify that, with inclusion of overlap, $e_i = \alpha - w\gamma/(1 - Sw)$, where $\gamma \equiv \beta - S\alpha$. (c) With S taken as 0.25, find the Wheland e_i values for benzene in terms of α and γ . (d) Show that the predicted

wavenumber of the transition between the highest-occupied (HO) and lowest-unoccupied (LU) Wheland π MOs is

$$\frac{1}{\lambda} = \frac{\gamma(w_{\text{HO}} - w_{\text{LU}})}{hc[1 + S^2 w_{\text{LU}} w_{\text{HO}} - S(w_{\text{LU}} + w_{\text{HO}})]}$$

Explain why $w_{\text{LU}} = -w_{\text{HO}}$ for an alternant hydrocarbon and use this relation to show that for an alternant hydrocarbon

$$1/\lambda = |\gamma| \Delta w / hc [1 - \frac{1}{4} S^2 (\Delta w)^2]$$

where $\Delta w \equiv |w_{\text{HO}} - w_{\text{LU}}|$. Δw is identical to Δx calculated without overlap; also, $S = 0.25$ and Δw is typically about 1 or 2, so $\frac{1}{4} S^2 (\Delta w)^2 \ll 1$. Hence the $1/\lambda$ value is nearly the same as calculated [Eq. (17.47)] without overlap, except that the empirical proportionality constant is interpreted as $|\gamma| = |\beta - S\alpha|$, rather than as $|\beta|$.

- 17.25** Verify (17.64) for the PPP F_π matrix elements.
- 17.26** Apply the extended Hückel method to H_2 . Use $K = 1.75$; use the expression (13.60) of S as a function of R ; plot the valence-electron energy as a function of R ; compare the predicted value of R_e with the experimental value. (To avoid solving a secular equation, use symmetry orbitals.) Use 1.0 as the orbital exponent.
- 17.27** (a) Set up (but do not solve) the 8×8 extended Hückel secular determinant for methane at the equilibrium configuration; do not use symmetry orbitals. Take $K = 1.75$; use Slater's rules (Prob. 15.62) for the orbital exponents; evaluate the overlap integrals from the reference of Prob. 15.29. (b) Do the same as in (a), but now use symmetry orbitals.
- 17.28** Let the valence AOs r_A, s_A, t_B, u_B, v_C be centered on different atoms A, B, and C in a molecule, with r_A and s_A being different AOs. Which of the following integrals are neglected in the CNDO method? Which are neglected in the INDO method? Which are neglected in the MNDO method? Which are neglected in the AM1 method? (a) $(r_A r_A | s_A s_A)$; (b) $(r_A s_A | r_A s_A)$; (c) $(r_A r_A | t_B t_B)$; (d) $(r_A t_B | r_A t_B)$; (e) $(r_A r_A | r_A t_B)$; (f) $(r_A s_A | t_B u_B)$; (g) $(r_A r_A | t_B v_C)$.
- 17.29** The AM1 valence electronic energies of the atoms H and O are -11.396 eV and -316.100 eV, respectively. For H_2O at its AM1-calculated equilibrium geometry, the AM1 valence electronic energy (core-core repulsion omitted) is -493.358 eV and the AM1 core-core repulsion energy is 144.796 eV. For H(g) and O(g) , $\Delta H_{f,298}^\circ$ values are 52.102 and 59.559 kcal/mol, respectively. Find the AM1 prediction of $\Delta H_{f,298}^\circ$ of $\text{H}_2\text{O(g)}$. The experimental value is -57.796 kcal/mol.
- 17.30** For each of the following molecules, do AM1 and PM3 calculations to find the predicted geometry, dipole moment, and gas-phase $\Delta H_{f,298}^\circ$. Compare with experimental $\Delta H_{f,298}^\circ$ values in thermodynamics tables. (a) $\text{CH}_3\text{CH}_2\text{CH}_3$; (b) H_2S ; (c) benzene.
- 17.31** Calculate the barrier to internal rotation in ethane using AM1 and PM3. Compare the results to experiment and to the HF/6-31G** result (Prob. 15.53).
- 17.32** Find the AM1 geometry, dipole moment, and vibrational frequencies of formaldehyde. Compare the results with experiment and with the HF/3-21G and 6-31G* results (Prob. 15.45).
- 17.33** Find the AM1 geometries and dipole moments of the gauche and anti conformers of butane and compare with HF/6-31G* results (Prob. 15.58). Also find the AM1 difference in $\Delta H_{f,298}^\circ$ for these conformers.
- 17.34** Find the AM1 geometries and dipole moments of the two conformers of formic acid and compare with HF results (Prob. 15.57). Also find the AM1 difference in $\Delta H_{f,298}^\circ$ for these conformers.
- 17.35** Use a program such as Spartan or the online WebMO Demo Server that provides animations of calculated vibrational normal modes to: (a) Find the PM6 normal-mode harmonic vibrational wavenumbers of CH_3OH . (Note: The program output will likely call these "frequencies," but the units of cm^{-1} means they are actually wavenumbers.) (b) Calculate the PM6 zero-point energy of methanol per molecule and per mole. (c) View the animation of each normal mode and state the wavenumber of each of the following vibrations: CO stretching, OH stretching, torsion (twisting) about the CO bond, COH bending, symmetric CH stretching. (You may

- want to rotate the molecule for a clearer view of the vibration.) (d) As noted in Sec. 15.12, the accuracy of computed vibrational frequencies is improved by use of a scale factor. For PM6 the scale factor for harmonic vibrational frequencies is 1.093 (comp.chem.umn.edu/freqscale). Multiply the wavenumbers in (a) by this scale factor and compare the results with the experimental methanol harmonic frequencies found in the literature. (If you have trouble finding experimental harmonic frequencies, use fundamental frequencies instead.) Then repeat part (a) with another semiempirical method and see how much the wavenumbers differ.
- 17.36** Find the AM1 geometries of the reactant, product, and transition state in the reaction $\text{HCN} \rightarrow \text{HNC}$ and compare with the HF/6-31G* results (Prob. 16.45).
- 17.37** Give the number of bond-stretching terms, bond-bending terms, bond-torsion terms, van der Waals terms, and electrostatic terms in a force field for each of the following: (a) CF_3OH ; (b) $\text{CCl}_3\text{CCl}_2\text{OH}$.
- 17.38** Equation (17.89) for the van der Waals interaction between two atoms has the form $V = a/R^{12} - b/R^6$. (a) If σ is the finite value of R at which V is zero, show that $V = b\sigma^6/R^{12} - b/R^6$. (b) If R^* is the value of R at the minimum in V , show that $R^* = 2^{1/6}\sigma$. (c) If $\epsilon = V(\infty) - V(R^*)$ is the depth of the van der Waals well, show that $b = 4\sigma^6\epsilon$. (d) Use these results to show that the van der Waals V can be written in each of the two forms in (17.89).
- 17.39** Use whatever molecular-mechanics program is available to you to calculate the geometries of the two conformers of HCOOH and find the energy difference between them. Omit zero-point vibrational energies and thermal energies in this and the following MM problems.
- 17.40** Use a molecular-mechanics program to calculate the barrier to internal rotation in ethane.
- 17.41** Use a molecular-mechanics program to find the geometries of the gauche and anti conformations of butane and the energy difference between them.
- 17.42** Repeat the vibrational wavenumber calculations of Prob. 17.35 (a), (b), and (c) using the MM3 force field and compare with the PM6 results. Assume the MM3 scale factor is 1.
- 17.43** Use a molecular-mechanics program to find the geometries of the two conformers of vinyl alcohol (CH_2CHOH) and the energy difference between them.
- 17.44** Use a molecular-mechanics program to find the geometries and the energy difference between (a) cis and trans 1,2-difluoroethylene; (b) cis and trans 1,2-dichloroethylene; (c) cis and trans 1,2-diiodoethylene. Compare your results with available literature data as to which is the more stable isomer of each pair.
- 17.45** MM3 parameters in kcal/mol to be used with (17.90) to calculate $\Delta H_{f,298}^\circ$ values of saturated hydrocarbons are $a_{\text{CH}} = -4.590$ and $a_{\text{CC}} = 2.447$; in addition, the following corrections (in kcal/mol) are included: 1.045 for each CH_3 group, -2.627 for each carbon bonded to three other carbons, -6.641 for each carbon bonded to four other carbons, 0.42 for each bond (except $\text{C}-\text{CH}_3$ bonds) with a low rotational barrier, -1.780 for each four-membered ring, and -5.508 for each five-membered ring. Given the following MM3 steric energies, calculate the gas-phase $\Delta H_{f,298}^\circ$ of each compound and compare with the experimental values given in parentheses: (a) 2.05 kcal/mol for propane (-24.82 kcal/mol); (b) 3.18 kcal/mol for isobutane (-32.15 kcal/mol); (c) 32.63 kcal/mol for cyclobutane (6.78 kcal/mol). (d) How many conformations must be considered when calculating the MM $\Delta H_{f,298}^\circ$ of butane?
- 17.46** Consider the thermal reaction $1,3,5\text{-hexatriene} \rightarrow 1,3\text{-cyclohexadiene}$. (a) Use the symmetry of the polyene HOMO to predict whether the reaction path is conrotatory or disrotatory. (The HMOs need not be found explicitly; all that is needed are the signs of the AOs in the HOMO, and these can be found from the pattern of nodes; the number of vertical nodal planes is zero for the ground state and increases by 1 for each higher state.) (b) Do the same as in (a) when the reaction occurs photochemically. (c) State the general rules for the cyclization of the polyene (17.28) with n_π π electrons.
- 17.47** Examine the frontier orbitals and decide whether each of the following elementary reactions should have a high or low activation energy for a four-center broadside-collision reaction path. (a) $\text{H}_2 + \text{D}_2 \rightarrow 2\text{HD}$; (b) $\text{N}_2 + \text{O}_2 \rightarrow 2\text{NO}$; (c) $\text{F}_2 + \text{Br}_2 \rightarrow 2\text{FBr}$; (d) $\text{H}_2 + \text{C}_2\text{H}_4 \rightarrow \text{C}_2\text{H}_6$.