CHAPTER TWO

An Introduction to Computational Quantum Mechanics

2.1 Introduction

Our aim in this chapter will be to establish the basic elements of those quantum mechanical methods that are most widely used in molecular modelling. We shall assume some familiarity with the elementary concepts of quantum mechanics as found in most 'general' physical chemistry textbooks, but little else other than some basic mathematics (see Section 1.10). There are also many excellent introductory texts to quantum mechanics. In Chapter 3 we then build upon this chapter and consider more advanced concepts. Quantum mechanics does, of course, predate the first computers by many years, and it is a tribute to the pioneers in the field that so many of the methods in common use today are based upon their efforts. The early applications were restricted to atomic, diatomic or highly symmetrical systems which could be solved by hand. The development of quantum mechanical techniques that are more generally applicable and that can be implemented on a computer (thereby eliminating the need for much laborious hand calculation) means that quantum mechanics can now be used to perform calculations on molecular systems of real, practical interest. Quantum mechanics explicitly represents the electrons in a calculation, and so it is possible to derive properties that depend upon the electronic distribution and, in particular, to investigate chemical reactions in which bonds are broken and formed. These qualities, which differentiate quantum mechanics from the empirical force field methods described in Chapter 4, will be emphasised in our discussion of typical applications.

There are a number of quantum theories for treating molecular systems. The first we shall examine, and the one which has been most widely used, is *molecular orbital theory*. However, alternative approaches have been developed, some of which we shall also describe, albeit briefly. We will be primarily concerned with the *ab initio* and semi-empirical approaches to quantum mechanics but will also mention techniques such as Hückel theory and valence bond theory. An alternative approach to quantum mechanics, density functional theory, is considered in Chapter 3. Density functional theory has always enjoyed significant support from the materials science community but is increasingly used for molecular systems.

Quantum mechanics is often considered to be a difficult subject, and a cursory glance at the following pages in this chapter may simply serve to reinforce that view! However, if followed carefully it is possible to see how models that are developed for very simple

systems can be applied to much more complex systems. As a consequence our treatment does require some consideration of the mathematical background to the simplest and most common types of calculation. Our strategy in developing the underlying theory of molecular orbital quantum mechanical calculations is as follows. First, we revise some key features of quantum mechanics, including the hydrogen atom. We then discuss the functional form of an acceptable wavefunction for a molecular system and show how to calculate the energy of such a system from the wavefunction. This leads to the problem of determining the wavefunction itself and how this can be done using routine mathematical methods. We will then be in a position to understand how quantum mechanical calculations can be performed for 'real' systems and will have the background necessary to consider more advanced topics.

The starting point for any discussion of quantum mechanics is, of course, the Schrödinger equation. The full, time-dependent form of this equation is

$$\left\{-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+\mathcal{V}\right\}\Psi(\mathbf{r},t)=i\hbar\frac{\partial\Psi(\mathbf{r},t)}{\partial t} \tag{2.1}$$

Equation (2.1) refers to a single particle (e.g. an electron) of mass m which is moving through space (given by a position vector $\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$) and time ($t$) under the influence of an external field $\mathcal{V}$ (which might be the electrostatic potential due to the nuclei of a molecule). $\hbar$ is Planck's constant divided by 2π and i is the square root of -1. Ψ is the *wavefunction* which characterises the particle's motion; it is from the wavefunction that we can derive various properties of the particle. When the external potential $\mathcal{V}$ is independent of time then the wavefunction can be written as the product of a spatial part and a time part: $\Psi(\mathbf{r},t)=\psi(\mathbf{r})T(t)$. We shall only consider situations where the potential is independent of time, which enables the time-dependent Schrödinger equation to be written in the more familiar, time-independent form:

$$\left\{-\frac{\hbar^2}{2m}\nabla^2+\mathcal{V}\right\}\Psi(\mathbf{r})=E\Psi(\mathbf{r}) \tag{2.2}$$

Here, E is the energy of the particle and we have used the abbreviation ∇^2 (pronounced 'del-squared').

$$\nabla^2=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2} \tag{2.3}$$

It is usual to abbreviate the left-hand side of Equation (2.1) to $\mathcal{H}\Psi$, where $\mathcal{H}$ is the *Hamiltonian operator*:

$$\mathcal{H}=-\frac{\hbar^2}{2m}\nabla^2+\mathcal{V} \tag{2.4}$$

This reduces the Schrödinger equation to $\mathcal{H}\Psi=E\Psi$. To solve the Schrödinger equation it is necessary to find values of E and functions Ψ such that, when the wavefunction is operated upon by the Hamiltonian, it returns the wavefunction multiplied by the energy. The Schrödinger equation falls into the category of equations known as partial differential eigenvalue equations in which an operator acts on a function (the eigenfunction) and returns the

function multiplied by a scalar (the eigenvalue). A simple example of an eigenvalue equation is:

$$\frac{d}{dx}(y) = ry \tag{2.5}$$

The operator here is d/dx. One eigenfunction of this equation is $y = e^{ax}$ with the eigenvalue r being equal to a. Equation (2.5) is a first-order differential equation. The Schrödinger equation is a second-order differential equation as it involves the second derivative of Ψ. A simple example of an equation of this type is

$$\frac{d^2y}{dx^2} = ry \tag{2.6}$$

The solutions of Equation (2.6) have the form $y = A\cos kx + B\sin kx$, where A, B and k are constants. In the Schrödinger equation Ψ is the eigenfunction and E the eigenvalue.

2.1.1 Operators

The concept of an operator is an important one in quantum mechanics. The *expectation value* (which we can consider to be the average value) of a quantity such as the energy, position or linear momentum can be determined using an appropriate operator. The most commonly used operator is that for the energy, which is the Hamiltonian operator itself, $\mathscr{H}$. The energy can be determined by calculating the following integral:

$$E = \frac{\int \Psi^* \mathscr{H} \Psi \, d\tau}{\int \Psi^* \Psi \, d\tau} \tag{2.7}$$

The two integrals in Equation (2.7) are performed over all space (i.e. from $-\infty$ to $+\infty$ in the x, y and z directions). Note the use of the complex conjugate notation (Ψ^*), which reminds us that the wavefunction may be a complex number. This equation can be derived by pre-multiplying both sides of the Schrödinger equation, $\mathscr{H}\Psi = E\Psi$, by the complex conjugate of the wavefunction, Ψ^*, and integrating both sides over all space. Thus:

$$\int \Psi^* \mathscr{H} \Psi \, d\tau = \int \Psi^* E \Psi \, d\tau \tag{2.8}$$

E is a scalar and so can be taken outside the integral, thus leading to Equation (2.7). If the wavefunction is normalised then the denominator in Equation (2.7) will equal 1.

The Hamiltonian operator is composed of two parts that reflect the contributions of kinetic and potential energies to the total energy. The kinetic energy operator is

$$-\frac{\hbar^2}{2m}\nabla^2 \tag{2.9}$$

and the operator for the potential energy simply involves multiplication by the appropriate expression for the potential energy. For an electron in an isolated atom or molecule the potential energy operator comprises the electrostatic interactions between the electron and the nucleus and the interactions between the electron and the other electrons For a

single electron and a single nucleus with Z protons the potential energy operator is thus:

$$\mathscr{V} = -\frac{Ze^2}{4\pi\varepsilon_0 r} \tag{2.10}$$

Another operator is that for linear momentum along the x direction, which is

$$\frac{\hbar}{i}\frac{\partial}{\partial x} \tag{2.11}$$

The expectation value of this quantity can thus be obtained by evaluating the following integral:

$$p_x = \frac{\int \Psi^* \frac{\hbar}{i}\frac{\partial}{\partial x}\Psi \, d\tau}{\int \Psi^* \Psi \, d\tau} \tag{2.12}$$

2.1.2 Atomic Units

Quantum mechanics is primarily concerned with atomic particles: electrons, protons and neutrons. When the properties of such particles (e.g. mass, charge, etc.) are expressed in 'macroscopic' units then the value must usually be multiplied or divided by several powers of 10. It is preferable to use a set of units that enables the results of a calculation to be reported as 'easily manageable' values. One way to achieve this would be to multiply each number by an appropriate power of 10. However, further simplification can be achieved by recognising that it is often necessary to carry quantities such as the mass of the electron or electronic charge all the way through a calculation. These quantities are thus also incorporated into the atomic units. The atomic units of length, mass and energy are as follows:

1 unit of charge equals the absolute charge on an electron, $|e| = 1.602\,19 \times 10^{-19}$ C
1 mass unit equals the mass of the electron, $m_e = 9.105\,93 \times 10^{-31}$ kg
1 unit of length (1 Bohr) is given by $a_0 = h^2/4\pi^2 m_e e^2 = 5.291\,77 \times 10^{-11}$ m
1 unit of energy (1 Hartree) is given by $E_a = e^2/4\pi\varepsilon_0 a_0 = 4.359\,81 \times 10^{-18}$ J

The atomic unit of length is the radius of the first orbit in Bohr's treatment of the hydrogen atom. It also turns out to be the most probable distance of a 1s electron from the nucleus in the hydrogen atom. The atomic unit of energy corresponds to the interaction between two electronic charges separated by the Bohr radius. The total energy of the 1s electron in the hydrogen atom equals −0.5 Hartree. In atomic units Planck's constant $h = 2\pi$ and so $\hbar \equiv 1$.

2.1.3 Exact Solutions to the Schrödinger Equation

The Schrödinger equation can be solved exactly for only a few problems, such as the particle in a box, the harmonic oscillator, the particle on a ring, the particle on a sphere and the hydrogen atom, all of which are dealt with in introductory textbooks. A common feature of these problems is that it is necessary to impose certain requirements (often called *boundary*

conditions) on possible solutions to the equation. Thus, for a particle in a box with infinitely high walls, the wavefunction is required to go to zero at the boundaries. For a particle on a ring the wavefunction must have a periodicity of 2π because it must repeat every traversal of the ring. An additional requirement on solutions to the Schrödinger equation is that the wavefunction at a point r, when multiplied by its complex conjugate, is the probability of finding the particle at the point (this is the Born interpretation of the wavefunction). The square of an electronic wavefunction thus gives the electron density at any given point. If we integrate the probability of finding the particle over all space, then the result must be 1 as the particle must be somewhere:

$$\int \Psi^* \Psi \, d\tau = 1 \tag{2.13}$$

$d\tau$ indicates that the integration is over all space. Wavefunctions which satisfy this condition are said to be *normalised*. It is usual to require the solutions to the Schrödinger equation to be orthogonal:

$$\int \Psi_m^* \Psi_n \, d\tau = 0 \quad (m \neq n) \tag{2.14}$$

A convenient way to express both the orthogonality of different wavefunctions and the normalisation conditions uses the *Kronecker delta*:

$$\int \Psi_m^* \Psi_n \, d\tau = \delta_{mn} \tag{2.15}$$

When used in this context, the Kronecker delta can be taken to have a value of 1 if m equals n and zero otherwise. Wavefunctions that are both orthogonal and normalised are said to be *orthonormal*.

2.2 One-electron Atoms

In an atom that contains a single electron, the potential energy depends upon the distance between the electron and the nucleus as given by the Coulomb equation. The Hamiltonian thus takes the following form:

$$\mathscr{H} = -\frac{\hbar^2}{2m}\nabla^2 - \frac{Ze^2}{4\pi\varepsilon_0 r} \tag{2.16}$$

In atomic units the Hamiltonian is:

$$\mathscr{H} = -\frac{1}{2}\nabla^2 - \frac{Z}{r} \tag{2.17}$$

For the hydrogen atom, the nuclear charge, Z, equals +1. r is the distance of the electron from the nucleus. The helium cation, He^+, is also a one-electron atom but has a nuclear charge of +2. As atoms have spherical symmetry it is more convenient to transform the Schrödinger equation to polar coordinates r, θ and ϕ, where r is the distance from the nucleus (located at the origin), θ is the angle to the z axis and ϕ is the angle from the x axis in the xy plane (Figure 2.1). The solutions can be written as the product of a radial function $R(r)$, which depends only on r, and an angular function $Y(\theta, \phi)$ called a *spherical harmonic*, which

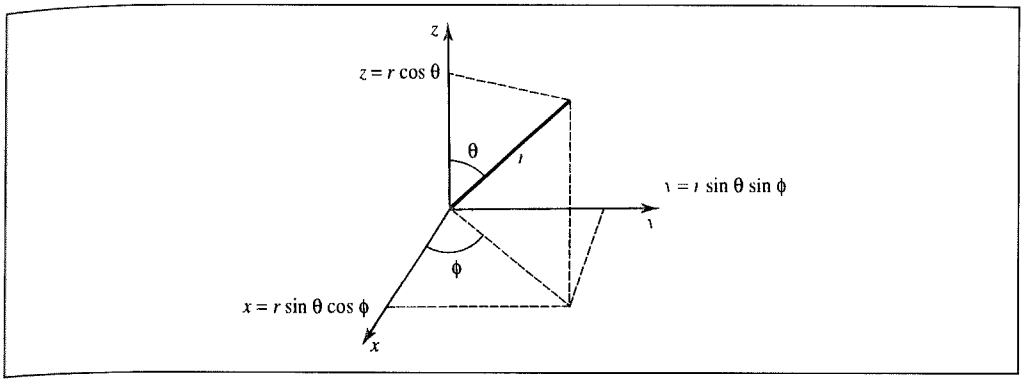

Fig. 2.1: The relationship between spherical polar and Cartesian coordinates

depends on θ and ϕ:

$$\Psi_{nlm} = R_{nl}(r)Y_{lm}(\theta, \phi) \tag{2.18}$$

The wavefunctions are commonly referred to as *orbitals* and are characterised by three quantum numbers n, m and l. The quantum numbers can adopt values as follows:

n: principal quantum number: 0, 1, 2, ...
l: azimuthal quantum number: 0, 1, ... $(n-1)$
m: magnetic quantum number: $-l$, $-(l-1)$, ... 0 ... $(l-1)$, l.

The full radial function is:

$$R_{nl}(r) = -\left[\left(\frac{2Z}{na_0}\right)^3 \frac{(n-l-1)!}{2n[(n+l)!]^3}\right]^{1/2} \exp\left(-\frac{\rho}{2}\right)\rho^l L_{n+1}^{2l+1}(\rho) \tag{2.19}$$

$\rho = 2Zr/na_0$, where a_0 is the Bohr radius.* The term in square brackets is a normalising factor. $L_{n+1}^{2l+1}(\rho)$ is a special type of function called a Laguerre polynomial. We shall rarely be interested in any other than the first few members of the series; moreover, they simplify considerably if atomic units are used and we write them in terms of the *orbital exponent* $\zeta = Z/n$. The first few members of the series for low values of n are given in Table 2.1 and are illustrated graphically in Figure 2.2. As can be seen, the radial part of the wavefunction is a polynomial multiplied by a decaying exponential.

The angular part of the wavefunction is the product of a function of θ and a function of ϕ:

$$Y_{lm}(\theta, \phi) = \Theta_{lm}(\theta)\Phi_m(\phi) \tag{2.20}$$

These functions are:

$$\Phi_m(\phi) = \frac{1}{\sqrt{2\pi}}\exp(im\phi) \tag{2.21}$$

$$\Theta_{lm}(\theta) = \left[\frac{(2l+1)}{2}\frac{(l-|m|)!}{(l+|m|)!}\right]^{1/2} P_l^{|m|}(\cos\theta) \tag{2.22}$$

*Strictly, a_0 in this case is given by $a_0 = h^2/\pi^2\mu e$, where μ is the reduced mass, $\mu = m_e M/(m_e + M)$, M is the mass of the nucleus.

n	l	$R_{nl}(r)$
1	0	$2\zeta^{3/2}\exp(-\zeta r)$
2	0	$2\zeta^{3/2}(1-\zeta r)\exp(-\zeta r)$
2	1	$(4/3)^{1/2}\zeta^{5/2}r\exp(-\zeta r)$
3	0	$(2/3)^{1/2}\zeta^{3/2}(3-6\zeta r+2\zeta^2 r^2)\exp(-\zeta r)$
3	1	$(8/9)^{1/2}\zeta^{5/2}(2-\zeta r)r\exp(-\zeta r)$
3	2	$(8/45)^{1/2}\zeta^{7/2}r^2\exp(-\zeta r)$

Table 2 1 Radial function for one-electron atoms.

The functions $\Phi_m(\phi)$ are just the solutions to the Schrödinger equation for a particle on a ring. The term in square brackets for the function $\Theta_{lm}(\theta)$ is a normalising factor. $P_l^{|m|}(\cos\theta)$ is a member of a series of functions called the associated Legendre polynomials (the 'Legendre polynomials' are functions for which $|m| = 0$). The total orbital angular momentum of an electron in the orbital is given by $l(l+1)\hbar$ and the component of the angular momentum along the $\theta = 0$ axis is given by $l\hbar$. The energy of each solution is a function of the principal quantum number only; thus orbitals with the same value of n but different l and m are degenerate. The orbitals are often represented as shown in Figure 2.3. These graphical representations are not necessarily the same as the solutions given above. For example, the 'correct' solutions for the 2p orbitals comprise one real and two complex functions:

$$2p(+1) = \sqrt{3/4\pi}R(r)\sin\theta e^{i\phi} \tag{2 23}$$

$$2p(0) = \sqrt{3/4\pi}R(r)\cos\theta \tag{2.24}$$

$$2p(-1) = \sqrt{3/4\pi}R(r)\sin\theta e^{-i\phi} \tag{2.25}$$

$R(r)$ is the radial part of the wavefunction and $\sqrt{3/4\pi}$ is a normalisation factor for the angular part. The 2p(0) function is real and corresponds to the $2p_z$ orbital that is pictured in Figure 2.3. A linear combination of the two remaining 2p solutions is used to generate two 'real' 2p wavefunctions, making use of the relationship $\exp(i\phi) = \cos\phi + i\sin\phi$ (Section 1.10.4). These linear combinations are the $2p_x$ and $2p_y$ orbitals shown in Figure 2.3.

$$2p_x = 1/2[2p(+1) + 2p(-1)] = \sqrt{3/4\pi}R(r)\sin\theta\cos\phi \tag{2.26}$$

$$2p_y = -1/2[2p(+1) - 2p(-1)] = \sqrt{3/4\pi}R(r)\sin\theta\sin\phi \tag{2.27}$$

These linear combinations still have the same energy as the original complex wavefunctions. This is a general property of degenerate solutions of the Hamiltonian operator. The reason why they are labelled $2p_x$ and $2p_y$ is that in polar coordinates the Cartesian coordinates x, y and z have the same angular dependence as the orbitals in Figure 2.3:

$$x = r\sin\theta\cos\phi \tag{2.28}$$

$$y = r\sin\theta\sin\phi \tag{2.29}$$

$$z = r\cos\theta \tag{2 30}$$

The solutions of the Schrödinger equation are either real or occur in degenerate pairs. These pairs are complex conjugates that can then be combined to give energetically equivalent real solutions. It is only when dealing with certain types of operator that it is necessary to retain a

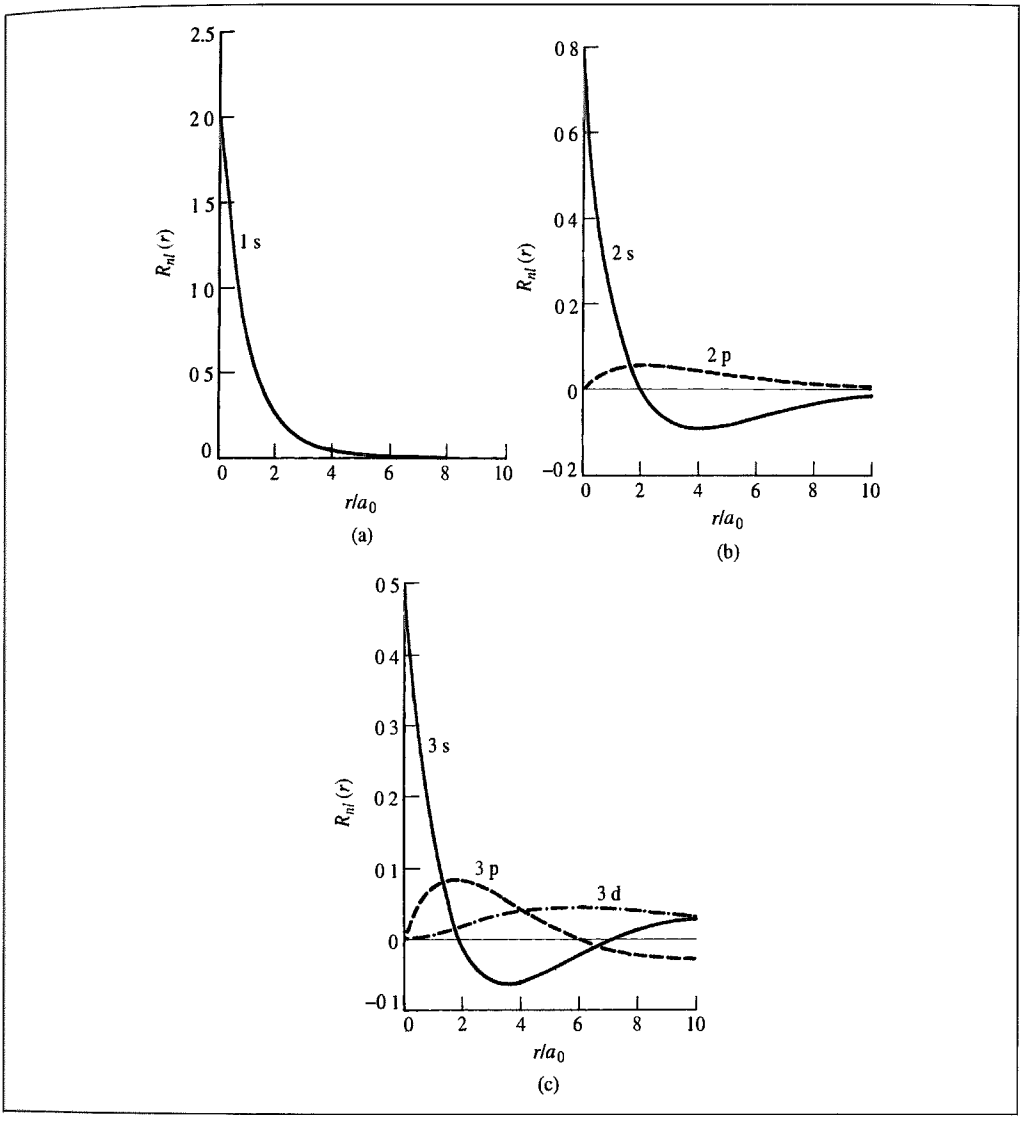

Fig. 2.2. The functions $R_{nl}(r)$ for the first three values of the principal quantum number. (a) 1s; (b) 2s and 2p; (c) 3s, 3p and 3d.

complex wavefunction (for the 2p functions, the operator that corresponds to angular momentum about the z axis falls into this category). In fact, to simplify matters we will almost always ignore the complex notation from now on and will deal with real orbitals.

Finally, we should note that the solutions are all orthogonal to each other; if the product of any pair of orbitals is integrated over all space, the result is zero unless the two orbitals are the same. Orthonormality is achieved by multiplying by an appropriate normalisation constant.

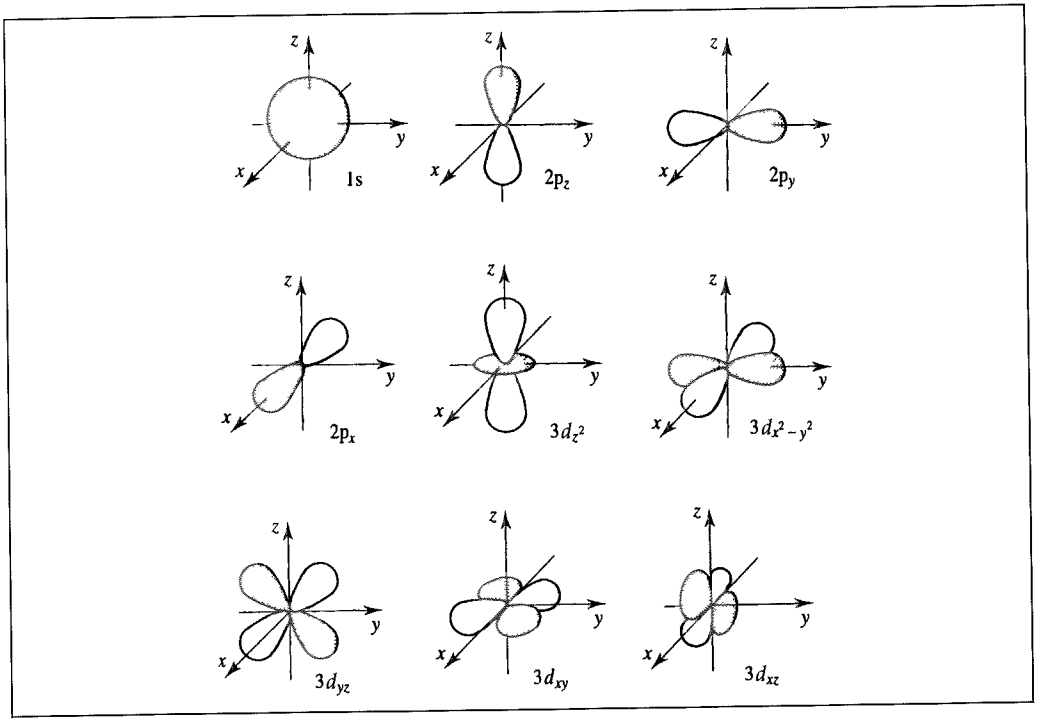

Fig 2 3: The common graphical representations of s, p and d orbitals

The orbital picture has proved invaluable for providing insight and qualitative interpretations into the nature of the bonding in and reactivity of chemical systems. It is one which we would like to retain for polyelectronic systems to provide a unifying theme that links the simplest systems with much more complicated ones.

2.3 Polyelectronic Atoms and Molecules

Solving the Schrödinger equation for atoms with more than one electron is complicated by a number of factors. The first complication is that the Schrödinger equation for such systems cannot be solved exactly, even for the helium atom. The helium atom has three particles (two electrons and one nucleus) and is an example of a *three-body problem*. No exact solutions can be found for systems that involve three (or more) interacting particles. Thus, any solutions we might find for polyelectronic atoms or molecules can only be approximations to the real, true solutions of the Schrödinger equation. One consequence of there being no exact solution is that the wavefunction may adopt more than one functional form; no form is necessarily more 'correct' than another. In fact, the most general form of the wavefunction will be an infinite series of functions.

A second complication with multi-electron species is that we must account for electron spin. Spin is characterised by the quantum number s, which for an electron can only take the

value $\frac{1}{2}$. The spin angular momentum is quantised such that its projection on the z axis is either $+\hbar$ or $-\hbar$. These two states are characterised by the quantum number m_s, which can have values of $+\frac{1}{2}$ or $-\frac{1}{2}$, and are often referred to as 'up spin' and 'down spin' respectively. Electron spin is incorporated into the solutions to the Schrödinger equation by writing each one-electron wavefunction as the product of a spatial function that depends on the coordinates of the electron and a spin function that depends on its spin. Such solutions are called *spin orbitals*, which we will represent using the symbol χ. The spatial part (which will be referred to as an orbital and represented using ϕ for atomic orbitals and ψ for molecular orbitals) describes the distribution of electron density in space and is analogous to the orbital diagrams in Figure 2.3. The spin part defines the electron spin and is labelled α or β. These spin functions have the value 0 or 1 depending on the quantum number m_s of the electron. Thus $\alpha(\frac{1}{2}) = 1$, $\alpha(-\frac{1}{2}) = 0$, $\beta(+\frac{1}{2}) = 0$, $\beta(-\frac{1}{2}) = 1$. Each spatial orbital can accommodate two electrons, with paired spins. In order to predict the electronic structure of a polyelectronic atom or a molecule, the *Aufbau principle* is employed, in which electrons are assigned to the orbitals, two electrons per orbital. We need to remember that electrons occupy degenerate states with a maximum number of unpaired electrons (Hund's rules), and that there are certain situations where it is energetically more favourable to place an unpaired electron in a higher-energy spatial orbital rather than pair it with another electron. However, such situations are rare, particularly for molecular systems, and for most of the situations that we shall be interested in the number of electrons, N, will be an even number that occupy the $N/2$ lowest-energy orbitals.

Electrons are indistinguishable. If we exchange any pair of electrons, then the distribution of electron density remains the same. According to the Born interpretation, the electron density is equal to the square of the wavefunction. It therefore follows that the wavefunction must either remain unchanged when two electrons are exchanged, or else it must change sign. In fact, for electrons the wavefunction is required to change sign: this is the *antisymmetry principle.*

2.3.1 The Born–Oppenheimer Approximation

It was stated above that the Schrödinger equation cannot be solved exactly for any molecular systems However, it is possible to solve the equation exactly for the simplest molecular species, H_2^+ (and isotopically equivalent species such as HD^+), when the motion of the electrons is decoupled from the motion of the nuclei in accordance with the Born–Oppenheimer approximation. The masses of the nuclei are much greater than the masses of the electrons (the resting mass of the lightest nucleus, the proton, is 1836 times heavier than the resting mass of the electron). This means that the electrons can adjust almost instantaneously to any changes in the positions of the nuclei. The electronic wavefunction thus depends only on the positions of the nuclei and not on their momenta. Under the Born–Oppenheimer approximation the total wavefunction for the molecule can be written in the following form:

$$\Psi_{\text{tot}}(\text{nuclei, electrons}) = \Psi(\text{electrons})\Psi(\text{nuclei}) \tag{2.31}$$

The total energy equals the sum of the nuclear energy (the electrostatic repulsion between the positively charged nuclei) and the electronic energy. The electronic energy comprises

the kinetic and potential energy of the electrons moving in the electrostatic field of the nuclei, together with electron–electron repulsion: $E_{tot} = E(\text{electrons}) + E(\text{nuclei})$.

When the Born–Oppenheimer approximation is used we concentrate on the electronic motions; the nuclei are considered to be fixed. For each arrangement of the nuclei the Schrödinger equation is solved for the electrons alone in the field of the nuclei. If it is desired to change the nuclear positions then it is necessary to add the nuclear repulsion to the electronic energy in order to calculate the total energy of the configuration.

2.3.2 The Helium Atom

We now return to the helium atom, our objective being to find a wavefunction that describes the behaviour of the electrons. The Born–Oppenheimer approximation is not, of course, relevant to systems with just one nucleus, and the wavefunction will be a function of the two electrons (which we shall label 1 and 2 with positions in space $\mathbf{r}_1$ and $\mathbf{r}_2$). As noted above, for polyelectronic systems any solution we find can only ever be an approximation to the true solution. There are a number of ways in which approximate solutions to the Schrödinger equation can be found. One approach is to find a simpler but related problem that can be more easily solved and then consider how the differences between the two problems change the Hamiltonian and thereby affect the solutions. This is called *perturbation theory* and is most appropriate when the differences between the real and simple problems are small. For example, a perturbation approach to tackling the helium atom might choose as the related system a 'pseudo atom', containing two electrons that interact with the nucleus but not with each other. Although this is a 'three-body' problem, the lack of any interaction between the electrons means that it can be solved exactly using the method of the separation of variables. The separation of variables technique can be applied whenever the Hamiltonian can be divided into parts that are themselves dependent solely upon subsets of the coordinates. The equation to be solved in this case is:

$$\left\{-\frac{\hbar^2}{2m}\nabla_1^2 - \frac{Ze^2}{4\pi\varepsilon_0 r_1} - \frac{\hbar^2}{2m}\nabla_2^2 - \frac{Ze^2}{4\pi\varepsilon_0 r_2}\right\}\Psi(\mathbf{r}_1,\mathbf{r}_2) = E\Psi(\mathbf{r}_1,\mathbf{r}_2) \tag{2.32}$$

Or, in atomic units,

$$\left\{-\frac{1}{2}\nabla_1^2 - \frac{Z}{r_1} - \frac{1}{2}\nabla_2^2 - \frac{Z}{r_2}\right\}\Psi(\mathbf{r}_1,\mathbf{r}_2) = E\Psi(\mathbf{r}_1,\mathbf{r}_2) \tag{2.33}$$

We can abbreviate this equation to

$$\{\mathscr{H}_1 + \mathscr{H}_2\}\Psi(\mathbf{r}_1,\mathbf{r}_2) = E\Psi(\mathbf{r}_1,\mathbf{r}_2) \tag{2.34}$$

$\mathscr{H}_1$ and $\mathscr{H}_2$ are the individual Hamiltonians for electrons 1 and 2. Let us assume that the wavefunction can be written as a product of individual one-electron wavefunctions, $\phi_1(\mathbf{r}_1)$ and $\phi_2(\mathbf{r}_2)$: $\Psi(\mathbf{r}_1,\mathbf{r}_2) = \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)$. Then we can write:

$$[\mathscr{H}_1 + \mathscr{H}_2]\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) = E\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) \tag{2.35}$$

Premultiplying by $\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)$ and integrating over all space gives:

$$\iint d\tau_1\, d\tau_2 \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)[\mathscr{H}_1 + \mathscr{H}_2]\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) = \iint d\tau_1\, d\tau_2 \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) \qquad (2.36)$$

or

$$\int d\tau_1 \phi_1(\mathbf{r}_1)\mathscr{H}_1\phi_1(\mathbf{r}_1) \int d\tau_2 \phi_2(\mathbf{r}_2)\phi_2(\mathbf{r}_2) + \int d\tau_1 \phi_1(\mathbf{r}_1)\phi_1(\mathbf{r}_1) \int d\tau_2 \phi_2(\mathbf{r}_2)\mathscr{H}_2\phi_2(\mathbf{r}_2)$$

$$= E \int d\tau_1 \phi_1(\mathbf{r}_1)\phi_1(\mathbf{r}_1) \int d\tau_2 \phi_2(\mathbf{r}_2)\phi_2(\mathbf{r}_2) \qquad (2.37)$$

If we assume that the wavefunctions are normalised then it can easily be seen that the total energy E is the sum of the individual orbital energies E_1 and E_2 ($E_1 = \int d\tau_1 \phi_1(\mathbf{r}_1)\mathscr{H}_1\phi_1(\mathbf{r}_1)$ and $E_2 = \int d\tau_2 \phi_2(\mathbf{r}_2)\mathscr{H}_2\phi_2(\mathbf{r}_2)$). When the separation of variables method is used the solutions for each electron are just those of the hydrogen atom (1s, 2s, etc.) in Equation (2.19) with $Z = 2$.

We now wish to establish the general functional form of possible wavefunctions for the two electrons in this pseudo helium atom. We will do so by considering first the spatial part of the wavefunction. We will show how to derive functional forms for the wavefunction in which the exchange of electrons is independent of the electron labels and does not affect the electron density. The simplest approach is to assume that each wavefunction for the helium atom is the product of the individual one-electron solutions. As we have just seen, this implies that the total energy is equal to the sum of the one-electron orbital energies, which is not correct as it ignores electron–electron repulsion. Nevertheless, it is a useful illustrative model. The wavefunction of the lowest energy state then has each of the two electrons in a 1s orbital:

$$1s(1)1s(2) \qquad (2.38)$$

'1s(1)' indicates a 1s function that depends on the coordinates of electron 1 ($\mathbf{r}_1$) and '1s(2)' indicates a 1s function that depends upon the coordinates of electron 2 ($\mathbf{r}_2$). This wavefunction satisfies the indistinguishability criterion, for we obtain the same function when we exchange the electrons – 1s(1)1s(2) is the same as 1s(2)1s(1). Its energy is twice that of a single electron in a 1s orbital. What of the first excited state, in which one electron is promoted to the 2s orbital? Two possible wavefunctions for this state are:

$$1s(1)2s(2) \qquad (2.39)$$

$$1s(2)2s(1) \qquad (2.40)$$

Do these wavefunctions satisfy the indistinguishability criterion? In other words, do we get the same function (or its negative) when we exchange the electrons? We do not, for when the two electrons (1 and 2) are exchanged then a different wavefunction is obtained: '1s(1)2s(2)' and '1s(2)2s(1)' are not the same, nor is one simply minus the other. However, linear combinations of these two wavefunctions do not suffer from the labelling problem and so we might anticipate that functional forms such as the following might constitute acceptable solutions to the Schrödinger equation for the pseudo helium atom:

$$(1/\sqrt{2})[1s(1)2s(2) + 1s(2)2s(1)] \qquad (2.41)$$

$$(1/\sqrt{2})[1s(1)2s(2) - 1s(2)2s(1)] \qquad (2.42)$$

The factor $(1/\sqrt{2})$ ensures that the wavefunction is normalised. Of the three acceptable spatial forms that we have described so far, two are symmetric (i.e. do not change sign when the electron labels are exchanged) and one is antisymmetric (the sign changes when the electrons are exchanged):

$$1s(1)1s(2) \quad \text{symmetric} \tag{2.43}$$

$$(1/\sqrt{2})[1s(1)2s(2) + 1s(2)2s(1)] \quad \text{symmetric} \tag{2.44}$$

$$(1/\sqrt{2})[1s(1)2s(2) - 1s(2)2s(1)] \quad \text{antisymmetric} \tag{2.45}$$

We now need to consider the effects of electron spin. For two electrons 1 and 2 there are four spin states; $\alpha(1)$, $\beta(1)$, $\alpha(2)$, $\beta(2)$. The indistinguishability criterion holds for the spin components as well, and so the following combinations of spin wavefunctions are possible.

$$\alpha(1)\alpha(2) \quad \text{symmetric} \tag{2.46}$$

$$\beta(1)\beta(2) \quad \text{symmetric} \tag{2.47}$$

$$(1/\sqrt{2})[\alpha(1)\beta(2) + \alpha(2)\beta(1)] \quad \text{symmetric} \tag{2.48}$$

$$(1/\sqrt{2})[\alpha(1)\beta(2) - \alpha(2)\beta(1)] \quad \text{antisymmetric} \tag{2.49}$$

When we combine the spatial and spin wavefunctions, the overall wavefunction must be antisymmetric with respect to exchange of electrons. It is therefore only admissible to combine a symmetric spatial part with an antisymmetric spin part, or an antisymmetric spatial part with a symmetric spin part. The following functional forms are therefore permissible functional forms for the wavefunctions of the ground and first few excited states of the helium atom:

$$(1/\sqrt{2})1s(1)1s(2)[\alpha(1)\beta(2) - \alpha(2)\beta(1)] \tag{2.50}$$

$$(1/2)[1s(1)2s(2) + 1s(2)2s(1)][\alpha(1)\beta(2) - \alpha(2)\beta(1)] \tag{2.51}$$

$$(1/\sqrt{2})[1s(1)2s(2) - 1s(2)2s(1)]\alpha(1)\alpha(2) \tag{2.52}$$

$$(1/\sqrt{2})[1s(1)2s(2) - 1s(2)2s(1)]\beta(1)\beta(2) \tag{2.53}$$

$$(1/2)[1s(1)2s(2) - 1s(2)2s(1)][\alpha(1)\beta(2) + \alpha(2)\beta(1)] \tag{2.54}$$

2.3.3 General Polyelectronic Systems and Slater Determinants

We now turn to the general case. What is an appropriate functional form of the wavefunction for a polyelectronic system (not necessarily an atom) with N electrons that satisfies the antisymmetry principle? First, we note that the following functional form of the wavefunction is inappropriate:

$$\Psi(1, 2, \ldots N) = \chi_1(1)\chi_2(2)\ldots\chi_N(N) \tag{2.55}$$

This product of spin orbitals is unacceptable because it does not satisfy the antisymmetry principle; exchanging pairs of electrons does not give the negative of the wavefunction. This formulation of the wavefunction is known as a *Hartree product*. The energy of a system described by a Hartree product equals the sum of the one-electron spin orbitals. A key conclusion of the Hartree product description is that the probability of finding an electron at a particular point in space is independent of the probability of finding any

other electron at that point in space. In fact, it turns out that the motions of the electrons are correlated. In addition, the Hartree product assumes that specific electrons have been assigned to specific orbitals, whereas the antisymmetry principle requires that the electrons are indistinguishable. Recall that for the helium atom, an acceptable functional form for the lowest-energy state, is:

$$\begin{aligned}\psi &= 1s(1)1s(2)[\alpha(1)\beta(2) - \alpha(2)\beta(1)] \\ &\equiv 1s(1)1s(2)\alpha(1)\beta(2) - 1s(1)1s(2)\alpha(2)\beta(1)\end{aligned} \tag{2.56}$$

This can be written in the form of a 2×2 determinant:

$$\begin{vmatrix} 1s(1)\alpha(1) & 1s(1)\beta(1) \\ 1s(2)\alpha(2) & 1s(2)\beta(2) \end{vmatrix} \tag{2.57}$$

The two spin orbitals are

$$\chi_1 = 1s(1)\alpha(1) \quad \text{and} \quad \chi_2 = 1s(1)\beta(1) \tag{2.58}$$

A determinant is the most convenient way to write down the permitted functional forms of a polyelectronic wavefunction that satisfies the antisymmetry principle. In general, if we have N electrons in spin orbitals $\chi_1, \chi_2, \ldots, \chi_N$ (where each spin orbital is the product of a spatial function and a spin function) then an acceptable form of the wavefunction is:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_N(2) \\ \vdots & \vdots & & \vdots \\ \chi_1(N) & \chi_2(N) & \cdots & \chi_N(N) \end{vmatrix} \tag{2.59}$$

As before, $\chi_1(1)$ is used to indicate a function that depends on the space and spin coordinates of the electron labelled '1'. The factor $1/\sqrt{N!}$ ensures that the wavefunction is normalised; we shall see later why the normalisation factor has this particular value. This functional form of the wavefunction is called a *Slater determinant* and is the simplest form of an orbital wavefunction that satisfies the antisymmetry principle. The Slater determinant is a particularly convenient and concise way to represent the wavefunction due to the special properties of determinants. Exchanging any two rows of a determinant, a process which corresponds to exchanging two electrons, changes the sign of the determinant and therefore directly leads to the antisymmetry property. If any two rows of a determinant are identical, which would correspond to two electrons being assigned to the same spin orbital, then the determinant vanishes. This can be considered a manifestation of the Pauli principle, which states that no two electrons can have the same set of quantum numbers. The Pauli principle also leads to the notion that each spatial orbital can accommodate two electrons of opposite spins.

When the Slater determinant is expanded, a total of $N!$ terms results. This is because there are $N!$ different permutations of N electrons. For example, for a three-electron system with spin orbitals χ_1, χ_2 and χ_3 the determinant is

$$\Psi = \frac{1}{\sqrt{12}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \chi_3(1) \\ \chi_1(2) & \chi_2(2) & \chi_3(2) \\ \chi_1(3) & \chi_2(3) & \chi_3(3) \end{vmatrix} \tag{2.60}$$

Expansion of the determinant gives the following expression (ignoring the normalisation constant):

$$\chi_1(1)\chi_2(2)\chi_3(3) - \chi_1(1)\chi_3(2)\chi_2(3) + \chi_2(1)\chi_3(2)\chi_1(3) - \chi_2(1)\chi_1(2)\chi_3(3) + \chi_3(1)\chi_1(2)\chi_2(3) - \chi_3(1)\chi_2(2)\chi_1(3) \quad (2.61)$$

This expansion contains six terms ($\equiv 3!$). The six possible permutations of three electrons are: 123, 132, 213, 231, 312, 321. Some of these permutations involve single exchanges of electrons; others involve the exchange of two electrons. For example, the permutation 132 can be generated from the initial permutation by exchanging electrons 2 and 3. If we do so then the following wavefunction is obtained:

$$\begin{aligned}
&\chi_1(1)\chi_2(3)\chi_3(2) - \chi_1(1)\chi_3(3)\chi_2(2) + \chi_2(1)\chi_3(3)\chi_1(2) \\
&\quad - \chi_2(1)\chi_1(3)\chi_3(2) + \chi_3(1)\chi_1(3)\chi_2(2) - \chi_3(1)\chi_2(3)\chi_1(2) \\
&= -\chi_1(1)\chi_2(2)\chi_3(3) + \chi_1(1)\chi_3(2)\chi_2(3) - \chi_2(1)\chi_3(2)\chi_1(3) \\
&\quad + \chi_2(1)\chi_1(2)\chi_3(3) - \chi_3(1)\chi_1(2)\chi_2(3) + \chi_3(1)\chi_2(2)\chi_1(3) \\
&= -\Psi
\end{aligned} \quad (2.62)$$

By contrast, the permutation 312 requires that electrons 1 and 3 are exchanged and then electrons 1 and 2 are exchanged. This gives rise to an unchanged wavefunction. In general, an odd permutation involves an odd number of electron exchanges and leads to a wavefunction with a changed sign; an even permutation involves an even number of electron exchanges and returns the wavefunction unchanged.

For any sizeable system the Slater determinant can be tedious to write out, let alone the equivalent full orbital expansion, and so it is common to use a shorthand notation. Various notation systems have been devised. In one system the terms along the diagonal of the matrix are written as a single-row determinant. For the 3×3 determinant we therefore have:

$$\begin{vmatrix} \chi_1(1) & \chi_2(1) & \chi_3(1) \\ \chi_1(2) & \chi_2(2) & \chi_3(2) \\ \chi_1(3) & \chi_2(3) & \chi_3(3) \end{vmatrix} \equiv |\chi_1 \quad \chi_2 \quad \chi_3| \quad (2.63)$$

The normalisation factor is assumed. It is often convenient to indicate the spin of each electron in the determinant; this is done by writing a bar when the spin part is β (spin down); a function without a bar indicates an α spin (spin up). Thus, the following are all commonly used ways to write the Slater determinantal wavefunction for the beryllium atom (which has the electronic configuration $1s^2 2s^2$):

$$\begin{aligned}
\Psi &= \frac{1}{\sqrt{24}} \begin{vmatrix} \phi_{1s}(1) & \bar{\phi}_{1s}(1) & \phi_{2s}(1) & \bar{\phi}_{2s}(1) \\ \phi_{1s}(2) & \bar{\phi}_{1s}(2) & \phi_{2s}(2) & \bar{\phi}_{2s}(2) \\ \phi_{1s}(3) & \bar{\phi}_{1s}(3) & \phi_{2s}(3) & \bar{\phi}_{2s}(3) \\ \phi_{1s}(4) & \bar{\phi}_{1s}(4) & \phi_{2s}(4) & \bar{\phi}_{2s}(4) \end{vmatrix} \\
&\equiv |\phi_{1s} \quad \bar{\phi}_{1s} \quad \phi_{2s} \quad \bar{\phi}_{2s}| \\
&\equiv |1s \quad \bar{1}s \quad 2s \quad \bar{2}s|
\end{aligned} \quad (2.64)$$

An important property of determinants is that a multiple of any column can be added to another column without altering the value of the determinant. This means that the spin orbitals are not unique; other linear combinations give the same energy. To illustrate this, consider the first excited state configuration of the helium atom ($1s^2 2s^2$), which can be written as the following 2×2 determinant:

$$\begin{vmatrix} 1s(1)\alpha(1) & 2s(1)\alpha(1) \\ 1s(2)\alpha(2) & 2s(2)\alpha(2) \end{vmatrix} = 1s(1)\alpha(1)2s(2)\alpha(2) - 1s(2)\alpha(2)2s(1)\alpha(1) \tag{2.65}$$

We now introduce two new 'spin orbitals':

$$\chi_1' = \frac{1s + 2s}{\sqrt{2}}\alpha; \quad \chi_2' = \frac{1s - 2s}{\sqrt{2}}\alpha \tag{2.66}$$

With these new orbitals the value of the determinant is as follows:

$$\begin{aligned} \begin{vmatrix} \chi_1'(1) & \chi_2'(1) \\ \chi_1'(2) & \chi_2'(2) \end{vmatrix} &= \frac{[1s(1) + 2s(1)][1s(2) - 2s(2)]\alpha(1)\alpha(2)}{2} \\ &\quad - \frac{[1s(1) - 2s(1)][1s(2) + 2s(2)]\alpha(1)\alpha(2)}{2} \\ &\equiv -\Psi \end{aligned} \tag{2.67}$$

This can be helpful because it may enable more meaningful sets of orbitals to be generated from the original solutions. Molecular orbital calculations may give solutions that are 'smeared out' throughout the entire molecule, whereas we may find orbitals that are localised in specific regions (e.g. in the bonds between atoms) to be more useful.

2.4 Molecular Orbital Calculations

2.4.1 Calculating the Energy from the Wavefunction: the Hydrogen Molecule

In our treatment of molecular systems we first show how to determine the energy for a given wavefunction, and then demonstrate how to calculate the wavefunction for a specific nuclear geometry. In the most popular kind of quantum mechanical calculations performed on molecules each molecular spin orbital is expressed as a linear combination of atomic orbitals (the LCAO approach*). Thus each molecular orbital can be written as a summation of the following form:

$$\psi_i = \sum_{\mu=1}^{K} c_{\mu i}\phi_\mu \tag{2.68}$$

ψ_i is a (spatial) molecular orbital, ϕ_μ is one of K atomic orbitals and $c_{\mu i}$ is a coefficient. In a simple LCAO picture of the lowest energy state of molecular hydrogen, H_2, there are two electrons with opposite spins in the lowest energy spatial orbital (labelled $1\sigma_g$), which is

* Computational quantum chemistry is well endowed with acronyms and abbreviations. A list of some of the more common ones can be found in Appendix 2.1

formed from a linear combination of two hydrogen-atom 1s orbitals:

$$1\sigma_g = A(1s_A + 1s_B) \tag{2.69}$$

A is the normalisation factor, whose value is not important in our present discussion. To calculate the energy of the ground state of the hydrogen molecule for a fixed internuclear distance we first write the wavefunction as a 2×2 determinant:

$$\Psi = \begin{vmatrix} \chi_1(1) & \chi_2(1) \\ \chi_1(2) & \chi_2(2) \end{vmatrix} = \chi_1(1)\chi_2(2) - \chi_1(2)\chi_2(1) \tag{2.70}$$

where

$$\begin{aligned} \chi_1(1) &= 1\sigma_g(1)\alpha(1) \\ \chi_2(1) &= 1\sigma_g(1)\beta(1) \\ \chi_1(2) &= 1\sigma_g(2)\alpha(2) \\ \chi_2(2) &= 1\sigma_g(2)\beta(2) \end{aligned} \tag{2.71}$$

For the hydrogen molecule, the Hamiltonian comprises the kinetic energy operator for each electron plus the potential energy operator due to the Coulomb attraction between the two electrons and the two nuclei, and the repulsion between the two electrons. In atomic units the Hamiltonian is thus

$$\mathscr{H} = -\frac{1}{2}\nabla_1^2 - \frac{1}{2}\nabla_2^2 - \frac{Z_A}{r_{1A}} - \frac{Z_B}{r_{1B}} - \frac{Z_A}{r_{2A}} - \frac{Z_B}{r_{2B}} + \frac{1}{r_{12}} \tag{2.72}$$

The electrons have been labelled 1 and 2 and the nuclei have been labelled A and B. For H_2 the nuclear charges Z_A and Z_B are both equal to 1. First we need to consider how to calculate the energy of this hydrogen molecule. This is obtained using Equation (2.7):

$$E = \frac{\int \Psi \mathscr{H} \Psi \, d\tau}{\int \Psi \Psi \, d\tau} \tag{2.73}$$

In general, a quantum mechanical calculation provides molecular orbitals that are normalised but the total wavefunction is not. The normalisation constant for the wavefunction of the two-electron hydrogen molecule is $1/\sqrt{2}$ and so the denominator in Equation (2.73) is equal to 2.

We now substitute the hydrogen molecule wavefunction into Equation (2.73) to provide the following:

$$\begin{aligned} E = \frac{1}{2}\iint d\tau_1 \, d\tau_2 \{[\chi_1(1)\chi_2(2) - \chi_2(1)\chi_1(2)][-\tfrac{1}{2}\nabla_1^2 - \tfrac{1}{2}\nabla_2^2 - (1/r_{1A}) - (1/r_{1B}) \\ - (1/r_{2A}) - (1/r_{2B}) + (1/r_{12})][\chi_1(1)\chi_2(2) - \chi_2(1)\chi_1(2)]\} \end{aligned} \tag{2.74}$$

$d\tau_i$ indicates that the integration is over the spatial and spin coordinates of electron i. It is useful to separate the Hamiltonian operator into two H_2^+ Hamiltonians plus the inter-electronic repulsion term:

$$\begin{aligned} E = \frac{1}{2}\iint d\tau_1 \, d\tau_2 \{[\chi_1(1)\chi_2(2) - \chi_2(1)\chi_1(2)][\mathscr{H}_1 + \mathscr{H}_2 + (1/r_{12})] \\ \times [\chi_1(1)\chi_2(2) - \chi_2(1)\chi_1(2)]\} \end{aligned} \tag{2.75}$$

where

$$\mathscr{H}_1 = -\frac{1}{2}\nabla_1^2 - \frac{1}{r_{1A}} - \frac{1}{r_{1B}} \quad \text{and} \quad \mathscr{H}_2 = -\frac{1}{2}\nabla_2^2 - \frac{1}{r_{2A}} - \frac{1}{r_{2B}} \tag{2.76}$$

We can now start to separate the integral in Equation (2.74) into individual terms and identify the various contributions to the electronic energy:

$$\begin{aligned} E = &\iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)(\mathscr{H}_1)\chi_1(1)\chi_2(2) \\ &- \iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)(\mathscr{H}_1)\chi_2(1)\chi_1(2) + \cdots \\ &+ \iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)(\mathscr{H}_2)\chi_1(1)\chi_2(2) \\ &- \iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)(\mathscr{H}_2)\chi_2(1)\chi_1(2) + \cdots \\ &+ \iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)\left(\frac{1}{r_{12}}\right)\chi_1(1)\chi_2(2) \\ &- \iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)\left(\frac{1}{r_{12}}\right)\chi_2(1)\chi_1(2) + \cdots \end{aligned} \tag{2.77}$$

Each of these individual terms can be simplified if we recognise that terms dependent upon electrons other than those in the operator can be separated out. For example, the first term in the expansion, Equation (2.77), is:

$$\iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)(\mathscr{H}_1)\chi_1(1)\chi_2(2) \tag{2.78}$$

The operator $\mathscr{H}_1$ is a function of the coordinates of electron 1 only, so terms involving electron 2 can be separated out as follows:

$$\begin{aligned} &\iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)(\mathscr{H}_1)\chi_1(1)\chi_2(2) \\ &\quad = \int d\tau_2 \chi_2(2)\chi_2(2) \int d\tau_1 \chi_1(1)\left(-\frac{1}{2}\nabla_1^2 - \frac{1}{r_{1A}} - \frac{1}{r_{1B}}\right)\chi_1(1) \end{aligned} \tag{2.79}$$

If the molecular orbitals are normalised, the integral $\int d\tau_2 \chi_2(2)\chi_2(2)$ equals 1. Further simplification can be achieved by splitting the integral involving electron 1 into separate integrals over the spatial and spin parts; the integral over spin orbitals is equal to the product of an integral over the spatial coordinates and an integral over the spin coordinates:

$$\begin{aligned} &\int d\tau_1 \chi_1(1)\left(-\nabla_1^2 - \frac{1}{r_{1A}} - \frac{1}{r_{1B}}\right)\chi_1(1) \\ &\quad = \int d\nu_1 1\sigma_g(1)\left(-\frac{1}{2}\nabla_1^2 - \frac{1}{r_{1A}} - \frac{1}{r_{1B}}\right)1\sigma_g(1) \int d\sigma_1 \alpha(1)\alpha(1) \end{aligned} \tag{2.80}$$

$d\nu$ indicates integration over spatial coordinates and $d\sigma$ indicates integration over the spin coordinates. The integral over the spin coordinates equals 1. This expression corresponds

to the sum of the kinetic and potential energy of an electron in the orbital $1\sigma_g$ in the electrostatic field of the two bare nuclei. This integral can in turn be expanded by substituting the atomic orbital combination for $1\sigma_g$:

$$\int d\nu_1 1\sigma_g(1)\left(-\frac{1}{2}\nabla_1^2 - \frac{1}{r_{1A}} - \frac{1}{r_{1B}}\right)1\sigma_g(1)$$
$$= A^2 \int d\nu_1 \{1s_A(1) + 1s_B(1)\}\left(-\frac{1}{2}\nabla_1^2 - \frac{1}{r_{1A}} - \frac{1}{r_{1B}}\right)\{1s_A(1) + 1s_B(1)\} \quad (2.81)$$

A is the normalisation constant. The integral in Equation (2.81) can in turn be factorised to give a sum of integrals, each of which involves a pair of atomic orbitals:

$$\int d\nu_1 \{1s_A(1) + 1s_B(1)\}\left(-\frac{1}{2}\nabla_1^2 - \frac{1}{r_{1A}} - \frac{1}{r_{1B}}\right)\{1s_A(1) + 1s_B(1)\}$$
$$= \int d\nu_1 1s_A(1)\left(-\frac{1}{2}\nabla_1^2 - \frac{1}{r_{1A}} - \frac{1}{r_{1B}}\right)1s_A(1)$$
$$+ \int d\nu_1 1s_A(1)\left(-\frac{1}{2}\nabla_1^2 - \frac{1}{r_{1A}} - \frac{1}{r_{1B}}\right)1s_B(1) + \cdots \quad (2.82)$$

Let us now apply the same procedure to the second term in Equation (2.77):

$$\iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)(\mathscr{H}_1)\chi_2(1)\chi_1(2) = \int d\tau_1 \chi_1(1)(\mathscr{H}_1)\chi_2(1) \int d\tau_2 \chi_2(2)\chi_1(2) \quad (2.83)$$

This particular integral is zero because the molecular orbitals are orthogonal and so the integral over the coordinates of electron 2 equals zero:

$$\int d\tau_2 \chi_2(2)\chi_1(2) = 0 \quad (2.84)$$

A similar procedure can be applied to the other integrals involving electron–nuclear interactions; it turns out that there are four non-zero integrals, each of which is equal to the energy of a single electron in the field of the two hydrogen nuclei.

There remain four integrals arising from electron–electron interactions. These are:

$$\iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)\left(\frac{1}{r_{12}}\right)\chi_1(1)\chi_2(2) + \iint d\tau_1\, d\tau_2 \chi_2(1)\chi_1(2)\left(\frac{1}{r_{12}}\right)\chi_2(1)\chi_1(2)$$
$$- \iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)\left(\frac{1}{r_{12}}\right)\chi_2(1)\chi_1(2) - \iint d\tau_1\, d\tau_2 \chi_2(1)\chi_1(2)\left(\frac{1}{r_{12}}\right)\chi_1(1)\chi_2(2) \quad (2.85)$$

The first two of these can be simplified as follows:

$$\iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)\left(\frac{1}{r_{12}}\right)\chi_1(1)\chi_2(2) = \iint d\nu_1\, d\nu_2 1\sigma_g(1)1\sigma_g(2)\left(\frac{1}{r_{12}}\right)1\sigma_g(1)1\sigma_g(2)$$
$$\times \int d\sigma_1 \alpha(1)\alpha(1) \int d\sigma_2 \beta(2)\beta(2)$$
$$= \iint d\nu_1\, d\nu_2 1\sigma_g(1)1\sigma_g(1)\left(\frac{1}{r_{12}}\right)1\sigma_g(2)1\sigma_g(2) \quad (2.86)$$

According to the Born interpretation of the wavefunction, $1\sigma_g(\mathbf{r}_1)1\sigma_g(\mathbf{r}_1)$ equals the electron density of electron 1 in orbital $1\sigma_g$ at a position $\mathbf{r}_1$. Similarly, $1\sigma_g(\mathbf{r}_2)1\sigma_g(\mathbf{r}_2)$ is the electron density of electron 2. The electrostatic repulsion between these regions of electron density thus equals $1\sigma_g(\mathbf{r}_1)1\sigma_g(\mathbf{r}_1) \times (1/r_{12}) \times 1\sigma_g(\mathbf{r}_2)1\sigma_g(\mathbf{r}_2)$, where r_{12} is the distance between the two electrons. The integral of this function over all space thus corresponds to the electrostatic (Coulomb) repulsion between the two orbitals.

If we substitute the atomic orbital expansion, we obtain a series of two-electron integrals, each of which involves four atomic orbitals:

$$\begin{aligned}\iint d\nu_1\, d\nu_2 1\sigma_g(1)1\sigma_g(2)\left(\frac{1}{r_{12}}\right)1\sigma_g(1)1\sigma_g(2) \\ = \iint d\nu_1\, d\nu_2 1s_A(1)1s_A(2)\left(\frac{1}{r_{12}}\right)1s_A(1)1s_A(2) \\ + \iint d\nu_1\, d\nu_2 1s_A(1)1s_A(2)\left(\frac{1}{r_{12}}\right)1s_A(1)1s_B(2) + \cdots\end{aligned} \tag{2.87}$$

The remaining two integrals from Equation (2.85) are:

$$\begin{aligned}\iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)\left(\frac{1}{r_{12}}\right)\chi_2(1)\chi_1(2) = \iint d\nu_1\, d\nu_2 1\sigma_g(1)1\sigma_g(2)\left(\frac{1}{r_{12}}\right)1\sigma_g(1)1\sigma_g(2) \\ \times \int d\sigma_1 \alpha(1)\beta(1) \int d\sigma_2 \beta(2)\alpha(2)\end{aligned} \tag{2.88}$$

$$\begin{aligned}\iint d\tau_1\, d\tau_2 \chi_2(1)\chi_1(2)\left(\frac{1}{r_{12}}\right)\chi_1(1)\chi_2(2) = \iint d\nu_1\, d\nu_2 1\sigma_g(1)1\sigma_g(2)\left(\frac{1}{r_{12}}\right)1\sigma_g(1)1\sigma_g(2) \\ \times \int d\sigma_1 \beta(1)\alpha(1) \int d\sigma_2 \alpha(2)\beta(2)\end{aligned} \tag{2.89}$$

Both of these integrals are zero due to the orthogonality of the electron spin states α and β.

The triplet excited state of H_2 is obtained by promoting an electron to a higher-energy molecular orbital. This higher-energy (antibonding) orbital is written $1\sigma_u$ and can be considered to arise from two 1s orbitals as follows:

$$1\sigma_u = A(1s_A - 1s_B) \tag{2.90}$$

The triplet state has two unpaired electrons with the same spin (α) and so the wavefunction state is:

$$\begin{vmatrix} 1\sigma_g\alpha(1) & 1\sigma_u\alpha(1) \\ 1\sigma_g\alpha(2) & 1\sigma_u\alpha(2) \end{vmatrix} \tag{2.91}$$

If we now expand the expression for the energy as for the ground state, terms analogous to the electron–nucleus and electron–electron interactions can again be obtained. However, the cross-terms are no longer equal to zero as was the case for the ground state, because the

electron spins are now the same (both α). For example, compare with Equation (2.88):

$$\iint d\tau_1\, d\tau_2 \chi_1(1)\chi_2(2)\left(\frac{1}{r_{12}}\right)\chi_2(1)\chi_1(2) = \iint d\nu_1\, d\nu_2 1\sigma_g(1)1\sigma_u(2)\left(\frac{1}{r_{12}}\right)1\sigma_g(2)1\sigma_u(1) \times \int d\sigma_1 \alpha(1)\alpha(1) \int d\sigma_2 \alpha(2)\alpha(2) \tag{2.92}$$

This contribution is called the *exchange interaction*. This appears with a minus sign in the expression for the total energy and so acts to stabilise the triplet $1s^1 2s^1$ state over the analogous singlet state. The exchange term is only non-zero for electrons of the same spin. It has the effect of making electrons of the same spin 'avoid' each other. As a result of this each electron can be considered to have a 'hole' associated with it. This hole is known as the *exchange hole* or the *Fermi hole*.

2.4.2 The Energy of a General Polyelectronic System

The hydrogen molecule is such a small problem that all of the integrals can be written out in full. This is rarely the case in molecular orbital calculations. Nevertheless, the same principles are used to determine the energy of a polyelectronic molecular system. For an N-electron system, the Hamiltonian takes the following general form:

$$\mathscr{H} = \left(-\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{r_{1A}} - \frac{1}{r_{1B}} \cdots + \frac{1}{r_{12}} + \frac{1}{r_{13}} + \cdots\right) \tag{2.93}$$

As with the hydrogen molecule, we have adopted the convention that the nuclei are labelled using capital letters A, B, C, etc., and the electrons are labelled 1, 2, 3, ...

Recall that the Slater determinant for a system of N electrons in N spin orbitals can be written:

$$\begin{vmatrix} \chi_1(1) & \chi_2(1) & \chi_3(1) & \cdots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \chi_3(2) & \cdots & \chi_N(2) \\ \chi_1(3) & \chi_2(3) & \chi_3(3) & \cdots & \chi_N(3) \\ \vdots & \vdots & \vdots & & \vdots \\ \chi_1(N) & \chi_2(N) & \chi_3(N) & \cdots & \chi_N(N) \end{vmatrix} \tag{2.94}$$

Each term in the determinant can thus be written $\chi_i(1)\chi_j(2)\chi_k(3)\ldots\chi_u(N-1)\chi_v(N)$ where i, j, k, ..., u, v is a series of N integers.

As usual, the energy can be calculated from $E = \int \Psi\mathscr{H}\Psi / \int \Psi\Psi$:

$$\int \Psi\mathscr{H}\Psi = \int\cdots\int d\tau_1\, d\tau_2 \cdots d\tau_N \Big\{ [\chi_i(1)\chi_j(2)\chi_k(3)\cdots] \times \left(-\frac{1}{2}\sum_i \nabla_i^2 - (1/r_{1A}) - (1/r_{1B})\cdots + (1/r_{12}) + (1/r_{13}) + \cdots\right) \times [\chi_i(1)\chi_j(2)\chi_k(3)\cdots]\Big\} \tag{2.95}$$

$$\int \Psi\Psi = \int \cdots \int d\tau_1\, d\tau_2 \cdots d\tau_N \{[\chi_i(1)\chi_j(2)\chi_k(3)\cdots][\chi_i(1)\chi_j(2)\chi_k(3)\cdots]\} \tag{2.96}$$

We can now see why the normalisation factor of the Slater determinantal wavefunction is $1/\sqrt{N!}$. If each determinant contains $N!$ terms then the product of two Slater determinants, [determinant][determinant], contains $(N!)^2$ terms. However, if the spin orbitals form an orthonormal set then only products of identical terms from the determinant will be non-zero when integrated over all space. We can illustrate this with the three-electron example. Considering just the first two terms in the expansion we obtain the following:

$$\begin{aligned}\iiint d\tau_1\, d\tau_2\, d\tau_3 [&\chi_1(1)\chi_2(2)\chi_3(3) - \chi_1(1)\chi_3(2)\chi_2(3) + \cdot\ \cdot] \\ &\times [\chi_1(1)\chi_2(2)\chi_3(3) - \chi_1(1)\chi_3(2)\chi_2(3) + \cdots]\end{aligned} \tag{2.97}$$

When multiplied out this gives:

$$\begin{aligned}&\iiint d\tau_1\, d\tau_2\, d\tau_3 [\chi_1(1)\chi_2(2)\chi_3(3)][\chi_1(1)\chi_2(2)\chi_3(3)] \\ &\quad - \iiint d\tau_1\, d\tau_2\, d\tau_3 [\chi_1(1)\chi_2(2)\chi_3(3)][\chi_1(1)\chi_3(2)\chi_2(3)] + \cdots \\ &\quad + \iiint d\tau_1\, d\tau_2\, d\tau_3 [\chi_1(1)\chi_3(2)\chi_2(3)][\chi_1(1)\chi_3(2)\chi_2(3)] + \cdots\end{aligned} \tag{2.98}$$

The first of the integrals in Equation (2.98) equals 1 (if the spin orbitals are normalised). The second term is zero because the terms involving both electrons 2 and 3 are different (for example, the integral $\int d\tau_2 \chi_2(2)\chi_3(2)$ will be zero due to the orthogonality of the spin orbitals χ_2 and χ_3). The third term in Equation (2.98) will be equal to 1, and so on. It turns out that there are $N!$ such non-zero terms. Thus if each individual term in the determinant is normalised, then:

$$\int \Psi\Psi = N! \tag{2.99}$$

Hence the normalisation factor for the determinantal wavefunction is $1/\sqrt{N!}$.

Turning now to the numerator in the energy expression (Equation (2.95)), this can be broken down into a series of one-electron and two-electron integrals, as for the hydrogen molecule. Each of these individual integrals has the general form:

$$\int \ldots \int d\tau_1\, d\tau_2 \ldots [term1] operator [term2] \tag{2 100}$$

[*term*1] and [*term*2] each represent one of the $N!$ terms in the Slater determinant. To simplify this integral, we first recognise that all spin orbitals involving an electron that does not appear in the operator can be taken outside the integral. For example, if the operator is $1/r_{1A}$, then all spin orbitals other than those that depend on the coordinates of electron 1 can be separated from the integral. The orthogonality of the spin orbitals means that the integral will be zero unless all indices involving these other electrons are the same in

[*term*1] and [*term*2]. Again, to use our three-electron system as an example:

$$\begin{aligned}
&\iiint d\tau_1\, d\tau_2\, d\tau_3[\chi_1(1)\chi_2(2)\chi_3(3)]\left(-\frac{1}{r_{1A}}\right)[\chi_1(1)\chi_2(2)\chi_3(3)] \\
&\quad = \iint d\tau_2\, d\tau_3[\chi_2(2)\chi_3(3)][\chi_2(2)\chi_3(3)]\int d\tau_1\chi_1(1)\left(-\frac{1}{r_{1A}}\right)\chi_1(1) \\
&\quad = \int d\tau_1\chi_1(1)\left(-\frac{1}{r_{1A}}\right)\chi_1(1)
\end{aligned} \tag{2.101}$$

But:

$$\begin{aligned}
&\iiint d\tau_1\, d\tau_2\, d\tau_3[\chi_1(1)\chi_2(2)\chi_3(3)]\left(-\frac{1}{r_{1A}}\right)[\chi_1(1)\chi_3(2)\chi_2(3)] \\
&\quad = \iint d\tau_2\, d\tau_3[\chi_2(2)\chi_3(3)][\chi_3(2)\chi_2(3)]\int d\tau_1\chi_1(1)\left(-\frac{1}{r_{1A}}\right)\chi_1(1) \\
&\quad = 0
\end{aligned} \tag{2.102}$$

For integrals that involve two-electron operators (i.e. $1/r_{ij}$), only those terms that do not involve the coordinates of the two electrons can be taken outside the integral. For example:

$$\begin{aligned}
&\iiint d\tau_1\, d\tau_2\, d\tau_3[\chi_1(1)\chi_2(2)\chi_3(3)]\left(\frac{1}{r_{12}}\right)[\chi_1(1)\chi_2(2)\chi_3(3)] \\
&\quad = \iint d\tau_1\, d\tau_2[\chi_1(1)\chi_2(2)]\left(\frac{1}{r_{12}}\right)[\chi_1(2)\chi_2(2)]\int d\tau_3\chi_3(3)\chi_3(3) \\
&\quad = \iint d\tau_1\, d\tau_2[\chi_1(1)\chi_2(2)]\left(\frac{1}{r_{12}}\right)[\chi_1(2)\chi_2(2)]
\end{aligned} \tag{2.103}$$

But:

$$\begin{aligned}
&\iiint d\tau_1\, d\tau_2\, d\tau_3[\chi_1(1)\chi_2(2)\chi_3(3)]\left(\frac{1}{r_{12}}\right)[\chi_1(1)\chi_3(2)\chi_2(3)] \\
&\quad = \iint d\tau_1\, d\tau_2[\chi_1(1)\chi_2(2)]\left(\frac{1}{r_{12}}\right)[\chi_1(2)\chi_3(2)]\int d\tau_3\chi_3(3)\chi_2(3) \\
&\quad = 0
\end{aligned} \tag{2.104}$$

As a consequence of these results, most of the individual integrals in the expansion will be zero. Nevertheless, it can be readily envisaged that there will still be an extremely large number of integrals to consider for all except the smallest problems. It is thus more convenient to write the energy expression in a concise form that recognises the three types of interaction that contribute to the total electronic energy of the system.

First, there is the kinetic and potential energy of each electron moving in the field of the nuclei. The energy associated with this contribution for the molecular orbital χ_i is often written H_{ii}^{core} and for M nuclei is given by:

$$H_{ii}^{\text{core}} = \int d\tau_1\chi_i(1)\left(-\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}}\right)\chi_i(1) \tag{2.105}$$

For N electrons in N molecular orbitals this contribution to the total energy is:

$$E_{\text{total}}^{\text{core}} = \sum_{i=1}^{N} \int d\tau_1 \chi_i(1) \left(-\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{1A}} \right) \chi_i(1) = \sum_{i=1}^{N} H_{ii}^{\text{core}} \tag{2.106}$$

Here we have followed convention and have used the label '1' wherever there is an integral involving the coordinates of a single electron, even though the actual electron may not be 'electron 1' Similarly, when it is necessary to consider two electrons then the labels 1 and 2 are conventionally employed. H_{ii}^{core} makes a favourable (i.e. negative) contribution to the electronic energy.

The second contribution to the energy arises from the electrostatic repulsion between pairs of electrons. This interaction depends on the electron–electron distance and, as we have seen, is calculated from integrals such as:

$$J_{ij} = \iint d\tau_1 \, d\tau_2 \chi_i(1)\chi_j(2) \left(\frac{1}{r_{12}} \right) \chi_i(1)\chi_j(2) \tag{2.107}$$

The symbol J_{ij} is often used to represent this Coulomb interaction between electrons in spin orbitals i and j, and is unfavourable (i.e. positive). The total electrostatic interaction between the electron in orbital χ_i and the other $N-1$ electrons is a sum of all such integrals, where the summation index j runs from 1 to N, excluding i:

$$\begin{aligned} E_i^{\text{Coulomb}} &= \sum_{j \neq i}^{N} \int d\tau_1 \, d\tau_2 \chi_i(1)\chi_j(2) \frac{1}{r_{12}} \chi_j(2)\chi_i(1) \\ &\equiv \sum_{j \neq i}^{N} \int d\tau_1 \, d\tau_2 \chi_i(1)\chi_i(1) \frac{1}{r_{12}} \chi_j(2)\chi_j(2) \end{aligned} \tag{2.108}$$

The total Coulomb contribution to the electronic energy of the system is obtained as a double summation over all electrons, taking care to count each interaction just once:

$$E_{\text{total}}^{\text{Coulomb}} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \int d\tau_1 \, d\tau_2 \chi_i(1)\chi_i(1) \frac{1}{r_{12}} \chi_j(2)\chi_j(2) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} J_{ij} \tag{2.109}$$

The third contribution to the energy is the exchange 'interaction'. This has no classical counterpart and arises because the motions of electrons with parallel spins are correlated: whereas there is a finite probability of finding two electrons with opposite (i.e. paired) spins at the same point in space, where the spins are the same then the probability is zero. This can be considered a manifestation of the Pauli principle, for if two electrons occupied the same region of space and had parallel spins then they could be considered to have the same set of quantum numbers. Electrons with the same spin thus tend to 'avoid' each other, and they experience a lower Coulombic repulsion, giving a lower (i.e. more favourable) energy. The exchange interaction involves integrals of the form:

$$K_{ij} = \iint d\tau_1 \, d\tau_2 \chi_i(1)\chi_j(2) \left(\frac{1}{r_{12}} \right) \chi_i(2)\chi_j(1) \tag{2.110}$$

This integral is only non-zero if the spins of the electrons in the spin orbitals χ_i and χ_j are the same. The energy due to exchange is often represented as K_{ij}. The exchange energy between the electron in spin orbital χ_i and the other $N-1$ electrons is:

$$E_i^{\text{exchange}} = \sum_{j \neq i}^{N} \iint d\tau_1 \, d\tau_2 \chi_i(1)\chi_j(2)\left(\frac{1}{r_{12}}\right)\chi_i(2)\chi_j(1) \tag{2.111}$$

The total exchange energy is calculated thus:

$$E_{\text{total}}^{\text{exchange}} = \sum_{i=1}^{N} \sum_{j'=i+1}^{N} \iint d\tau_1 \, d\tau_2 \chi_i(1)\chi_j(2)\left(\frac{1}{r_{12}}\right)\chi_i(2)\chi_j(1) = \sum_{j=1}^{N} \sum_{j'=i+1}^{N} K_{ij} \tag{2.112}$$

The prime on the counter j' indicates that the summation is only over electrons with the same spin as electron i.

2.4.3 Shorthand Representations of the One- and Two-electron Integrals

Various shorthand ways have been devised to represent the integrals involved in an electronic structure calculation. The two-electron integrals J_{ij} and K_{ij} are particularly long-winded to write out. In one scheme the Coulomb interaction J_{ij} is written as:

$$\left\langle \chi_i^* \chi_j^* \left| \frac{1}{r_{12}} \right| \chi_i \chi_j \right\rangle \tag{2.113}$$

In this notation the complex parts are written on the left-hand side and the real parts on the right. Sometimes the χ symbol is eliminated:

$$\left\langle ij \left| \frac{1}{r_{12}} \right| ij \right\rangle \tag{2.114}$$

The exchange integrals would be written:

$$\left\langle ij \left| \frac{1}{r_{12}} \right| ji \right\rangle \tag{2.115}$$

in this notation.

A notation that is widely used in the chemical literature writes the orbitals that are functions of electron 1 on the left-hand side (with the complex conjugate orbital first, if appropriate) and the orbitals that are functions of electron 2 on the right-hand side (again with the complex conjugate orbital first). In this notation, which is the one that we will adopt, the Coulomb integral is written $(ii|jj)$ and the exchange integral $(ij|ji)$. The one-electron integrals such as Equation (2.105) are written as follows:

$$\left(i \left| -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{1A}} \right| j \right) \equiv \int d\tau_1 \chi_i(1) \left(-\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{1A}} \right) \chi_j(1) \tag{2.116}$$

When calculating the total energy of the system, we should not forget the Coulomb interaction between the nuclei; this is constant within the Born–Oppenheimer approximation for a given spatial arrangement of nuclei. When it is desired to change the nuclear positions,

it is of course necessary to take the internuclear repulsion energy into account, which is calculated using the Coulomb equation:

$$\sum_{A=1}^{M}\sum_{B=A+1}^{M}\frac{Z_A Z_B}{R_{AB}} \tag{2.117}$$

2.4.4 The Energy of a Closed-shell System

In molecular modelling we are usually concerned with the ground states of molecules, most of which have closed-shell configurations. In a closed-shell system containing N electrons in $N/2$ orbitals, there are two spin orbitals associated with each spatial orbital ψ_i: $\psi_i\alpha$ and $\psi_i\beta$. The electronic energy of such a system can be calculated in a manner analogous to that for the hydrogen molecule. First, there is the energy of each electron moving in the field of the bare nuclei For an electron in a molecular orbital χ_i, this contributes an energy H_{ii}^{core}. If there are two electrons in the orbital then the energy is $2H_{ii}^{\text{core}}$ and for $N/2$ orbitals the total contribution to the energy will be:

$$\sum_{i=1}^{N/2} 2H_{ii}^{\text{core}} \tag{2.118}$$

If we consider the electron–electron terms, the interaction between each pair of orbitals ψ_i and ψ_j involves a total of four electrons. There are four ways in which two electrons in one orbital can interact in a Coulomb sense with two electrons in a second orbital, thus giving $4J_{ij}$. However, there are just two ways to obtain paired electrons from this arrangement, giving a total exchange contribution of $-2K_{ij}$. Finally, the Coulomb interaction between each pair of electrons in the same orbital must be included; there is no exchange interaction because the electrons have paired spins. The total energy is thus given as:

$$E = 2\sum_{i=1}^{N/2} H_{ii}^{\text{core}} + \sum_{i=1}^{N/2}\sum_{j=i+1}^{N/2}(4J_{ij} - 2K_{ij}) + \sum_{i=1}^{N/2} J_{ii} \tag{2.119}$$

A more concise form of this equation can be obtained if we recognise that $J_{ii} = K_{ii}$:

$$E = 2\sum_{i=1}^{N/2} H_{ii}^{\text{core}} + \sum_{i=1}^{N/2}\sum_{j=1}^{N/2}(2J_{ij} - K_{ij}) \tag{2.120}$$

2.5 The Hartree–Fock Equations

In our hydrogen molecule calculation in Section 2.4.1 the molecular orbitals were provided as input, but in most electronic structure calculations we are usually trying to calculate the molecular orbitals. How do we go about this? We must remember that for many-body problems there is no 'correct' solution; we therefore require some means to decide whether one proposed wavefunction is 'better' than another. Fortunately, the *variation theorem* provides us with a mechanism for answering this question. The theorem states that the

energy calculated from an approximation to the true wavefunction will always be greater than the true energy. Consequently, the better the wavefunction, the lower the energy. The 'best' wavefunction is obtained when the energy is a minimum. At a minimum, the first derivative of the energy, δE, will be zero. The Hartree–Fock equations are obtained by imposing this condition on the expression for the energy, subject to the constraint that the molecular orbitals remain orthonormal. The orthonormality condition is written in terms of the *overlap integral*, S_{ij}, between two orbitals i and j. Thus

$$S_{ij} = \int \chi_i \chi_j \, d\tau = \delta_{ij} \quad (\delta_{ij} \text{ is the Kronecker delta}) \tag{2.121}$$

This type of constrained minimisation problem can be tackled using the method of Lagrange multipliers. In this approach (see Section 1.10.5 for a brief introduction to Lagrange multipliers) the derivative of the function to be minimised is added to the derivatives of the constraint(s) multiplied by a constant called a Lagrange multiplier. The sum is then set equal to zero. If the Lagrange multiplier for each of the orthonormality conditions is written λ_{ij}, then:

$$\delta E + \delta \sum_i \sum_j \lambda_{ij} S_{ij} = 0 \tag{2.122}$$

In the Hartree–Fock equations the Lagrange multipliers are actually written $-2\varepsilon_{ij}$ to reflect the fact that they are related to the molecular orbital energies. The equation to be solved is thus:

$$\delta E - 2\delta \sum_i \sum_j \varepsilon_{ij} S_{ij} = 0 \tag{2.123}$$

We will not describe in detail how this equation is solved, as it is rather complicated. However, a qualitative picture is possible. The major difference between polyelectronic systems and systems with single electrons is the presence of interactions between the electrons, which, as we have seen, are expressed as Coulomb and exchange integrals. Suppose we are given the task of finding the 'best' (i.e. lowest energy) wavefunction for a polyelectronic system. We wish to retain the orbital picture of the system, in which single electrons are assigned to individual spin orbitals. The problem is to find a solution which simultaneously enables all the electronic motions to be taken into account, as a change in the spin orbital for one electron will influence the behaviour of an electron in another spin orbital due to the coupling of the electronic motions. We concentrate on a single electron in a spin orbital χ_i in the field of the nuclei and the other electrons in their (fixed) spin orbitals χ_j. The Hamiltonian operator for the electron in χ_i contains three terms appropriate to the three different contributions to the energy that were identified above (core, Coulomb, exchange). The result can be written as an integro-differential equation for χ_i that has the following form:

$$\left[-\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} \right] \chi_i(1) + \sum_{j \neq i} \left[\int d\tau_2 \chi_j(2) \chi_j(2) \frac{1}{r_{12}} \right] \chi_i(1)$$
$$- \sum_{j \neq i} \left[\int d\tau_2 \chi_j(2) \chi_i(2) \frac{1}{r_{12}} \right] \chi_i(1) = \sum_j \varepsilon_{ij} \chi_j(1) \tag{2.124}$$

This expression can be tidied up by introducing three operators that represent the contributions to the energy of the spin orbital χ_i in the 'frozen' system:

The core Hamiltonian operator, $\mathscr{H}^{\text{core}}(1)$:

$$\mathscr{H}^{\text{core}}(1) = -\frac{1}{2}\nabla_1^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{1A}} \tag{2.125}$$

In the absence of any interelectronic interactions this would be the only operator present, corresponding to the motion of a single electron moving in the field of the bare nuclei.

The Coulomb operator, $\mathscr{J}_j(1)$:

$$\mathscr{J}_j(1) = \int d\tau_2 \chi_j(2)\frac{1}{r_{12}}\chi_j(2) \tag{2.126}$$

This operator corresponds to the average potential due to an electron in χ_j.

The exchange operator $\mathscr{K}_j(1)$:

$$\mathscr{K}_j(1)\chi_i(1) = \left[\int d\tau_2 \chi_j(2)\frac{1}{r_{12}}\chi_i(2)\right]\chi_j(1) \tag{2.127}$$

The form of this operator is rather unusual, insofar as it must be defined in terms of its effect when acting on the spin orbital χ_i.

Equation (2.124) can thus be written:

$$\mathscr{H}^{\text{core}}(1)\chi_i(1) + \sum_{j\neq i}^{N}\mathscr{J}_j(1)\chi_i(1) - \sum_{j\neq i}^{N}\mathscr{K}_j(1)\chi_i(1) = \sum_j \varepsilon_{ij}\chi_j(1) \tag{2.128}$$

Making use of the fact that $\{\mathscr{J}_i(1) - \mathscr{K}_i(1)\}\chi_i(1) = 0$ leads to the following form:

$$\left[\mathscr{H}^{\text{core}}(1) + \sum_{j=1}^{N}\{\mathscr{J}_j(1) - \mathscr{K}_j(1)\}\right]\chi_i(1) = \sum_{j=1}^{N}\varepsilon_{ij}\chi_j(1) \tag{2.129}$$

Or, more simply:

$$\mathscr{f}_i\chi_i = \sum_j \varepsilon_{ij}\chi_j \tag{2.130}$$

$\mathscr{f}_i$ is called the *Fock operator*:

$$\mathscr{f}_i(1) = \mathscr{H}^{\text{core}}(1) + \sum_{j=1}^{N}\{\mathscr{J}_j(1) - \mathscr{K}_j(1)\} \tag{2.131}$$

For a closed-shell system, the Fock operator has the following form:

$$\mathscr{f}_i(1) = \mathscr{H}^{\text{core}}(1) + \sum_{j=1}^{N/2}\{2\mathscr{J}_j(1) - \mathscr{K}_j(1)\} \tag{2.132}$$

The Fock operator is an effective one-electron Hamiltonian for the electron in the polyelectronic system However, written in this form of Equation (2.130), the Hartree-Fock

equations do not seem to be particularly useful: on the left-hand side we have the Fock operator acting on the molecular orbital χ_i, but this returns, not the molecular orbital multiplied by a constant as in a normal eigenvalue equation, but rather a series of orbitals χ_j multiplied by some unknown constants ε_{ij}. This is because the solutions to the Hartree–Fock equations are not unique. We have already seen that the value of a determinant is unaffected when the multiple of any column is added to another column. If such a transformation is performed on the Slater determinant, then a different set of constants ε'_{ij} would be obtained with the spin orbitals χ'_i being linear combinations of the first set. Certain transformations give rise to localised orbitals, which are particularly useful for understanding the chemical nature of the system. These localised orbitals are no more 'correct' than a delocalised set. Fortunately, it is possible to manipulate Equations (2.130) mathematically so that the Lagrangian multipliers are zero unless the indices i and j are the same. The Hartree–Fock equations then take on the standard eigenvalue form:

$$\mathcal{F}_i\chi_i = \varepsilon_i\chi_i \tag{2.133}$$

Recall that in setting up these equations, each electron has been assumed to move in a 'fixed' field comprising the nuclei and the other electrons. This has important implications for the way in which we attempt to find a solution, for any solution that we might find by solving the equation for one electron will naturally affect the solutions for the other electrons in the system. The general strategy is called a *self-consistent field* (SCF) approach. One way to solve these equations is as follows. First, a set of trial solutions χ_i to the Hartree–Fock eigenvalue equations are obtained. These are used to calculate the Coulomb and exchange operators. The Hartree–Fock equations are solved, giving a second set of solutions χ_i, which are used in the next iteration. The SCF method thus gradually refines the individual electronic solutions that correspond to lower and lower total energies until the point is reached at which the results for all the electrons are unchanged, when they are said to be *self-consistent*.

2.5.1 Hartree–Fock Calculations for Atoms and Slater's Rules

The Hartree-Fock equations are usually solved in different ways for atoms and for molecules. For atoms, the equations can be solved numerically if it is assumed that the electron distribution is spherically symmetrical. However, these numerical solutions are not particularly useful. Fortunately, analytical approximations to these solutions, which are very similar to those obtained for the hydrogen atom, can be used with considerable success. These approximate analytical functions thus have the form:

$$\psi = R_{nl}(r)Y_{lm}(\theta, \phi) \tag{2.134}$$

Y is a spherical harmonic (as for the hydrogen atom) and R is a radial function. The radial functions obtained for the hydrogen atom cannot be used directly for polyelectronic atoms due to the screening of the nuclear charge by the inner shell electrons, but the hydrogen atom functions are acceptable if the orbital exponent is adjusted to account for the screening effect. Even so, the hydrogen atom functions are not particularly convenient to use in molecular orbital calculations due to their complicated functional form. Slater [Slater 1930] suggested

a simpler analytical form for the radial functions:

$$R_{nl}(r) = (2\zeta)^{n+1/2}[(2n)!]^{-1/2} r^{n-1} e^{-\zeta r} \tag{2.135}$$

These functions are universally known as *Slater type orbitals* (STOs) and are just the leading term in the appropriate Laguerre polynomials. The first three Slater functions are as follows.

$$R_{1s}(r) = 2\zeta^{3/2} e^{-\zeta r} \tag{2.136}$$

$$R_{2s}(r) = R_{2p}(r) = \left(\frac{4\zeta^5}{3}\right)^{1/2} r e^{-\zeta r} \tag{2.137}$$

$$R_{3s}(r) = R_{3p}(r) = R_{3d}(r) = \left(\frac{8\zeta^7}{45}\right)^{1/2} r^2 e^{-\zeta r} \tag{2.138}$$

To obtain the whole orbital we must multiply $R(r)$ by the appropriate angular part. For example, we would use the following expressions for the 1s, 2s and $2p_z$ orbitals:

$$\phi_{1s}(\mathbf{r}) = \sqrt{\zeta^3/\pi}\exp(-\zeta r) \tag{2.139}$$

$$\phi_{2s}(\mathbf{r}) = \sqrt{\zeta^5/3\pi}\,\mathbf{r}\exp(-\zeta r) \tag{2.140}$$

$$\phi_{2p_z}(\mathbf{r}) = \sqrt{\zeta^5/\pi}\exp(-\zeta r)\cos\theta \tag{2.141}$$

Slater provided a series of empirical rules for choosing the orbital exponents ζ, which are given by:

$$\zeta = \frac{Z - \sigma}{n^*} \tag{2.142}$$

Z is the atomic number and σ is a *shielding constant*, determined as below. n^* is an effective principal quantum number, which takes the same value as the true principal quantum number for $n = 1$, 2 or 3, but for $n = 4$, 5, 6 has the values 3.7, 4.0, 4.2, respectively. The shielding constant is obtained as follows:

First, divide the orbitals into the following groups:

$$\text{(1s); (2s, 2p); (3s, 3p); (3d); (4s, 4p); (4d); (4f); (5s, 5p); (5d)} \tag{2.143}$$

For a given orbital, σ is obtained by adding together the following contributions:

(a) zero from an orbital further from the nucleus than those in the group;
(b) 0.35 from each other electron in the same group, but if the other orbital is the 1s then the contribution is 0.3;
(c) 1.0 for each electron in a group with a principal quantum number 2 or more fewer than the current orbital;
(d) for each electron with a principal quantum number 1 fewer than the current orbital: 1.0 if the current orbital is d or f; 0.85 if the current orbital is s or p.

The shielding constant for the valence electrons of silicon is obtained using Slater's rules as follows. The electronic configuration of Si is $(1s^2)(2s^22p^6)(3s^23p^2)$. We therefore count

3×0.35 under rule (b), 2.0 under rule (c) and 8×0.85 under rule (d), giving a total of 9.85. When subtracted from the atomic number (14) this gives 4.15 for the value of $Z - \sigma$.

2.5.2 Linear Combination of Atomic Orbitals (LCAO) in Hartree–Fock Theory

Direct solution of the Hartree–Fock equations is not a practical proposition for molecules and so it is necessary to adopt an alternative approach. The most popular strategy is to write each spin orbital as a linear combination of single electron orbitals:

$$\psi_i = \sum_{\nu=1}^{K} c_{\nu i}\phi_\nu \tag{2.144}$$

The one-electron orbitals ϕ_ν are commonly called *basis functions* and often correspond to the atomic orbitals. We will label the basis functions with the Greek letters μ, ν, λ and σ. In the case of Equation (2.144) there are K basis functions and we should therefore expect to derive a total of K molecular orbitals (although not all of these will necessarily be occupied by electrons). The smallest number of basis functions for a molecular system will be that which can just accommodate all the electrons in the molecule. More sophisticated calculations use more basis functions than a minimal set. At the *Hartree–Fock limit* the energy of the system can be reduced no further by the addition of any more basis functions; however, it may be possible to lower the energy below the Hartree–Fock limit by using a functional form of the wavefunction that is more extensive than the single Slater determinant.

In accordance with the variation theorem we require the set of coefficients $c_{\nu i}$ that gives the lowest-energy wavefunction, and some scheme for changing the coefficients to derive that wavefunction. For a given basis set and a given functional form of the wavefunction (i.e. a Slater determinant) the best set of coefficients is that for which the energy is a minimum, at which point

$$\frac{\partial E}{\partial c_{\nu i}} = 0 \tag{2.145}$$

for all coefficients $c_{\nu i}$. The objective is thus to determine the set of coefficients that gives the lowest energy for the system.

2.5.3 Closed-shell Systems and the Roothaan–Hall Equations

We shall initially consider a closed-shell system with N electrons in $N/2$ orbitals. The derivation of the Hartree–Fock equations for such a system was first proposed by Roothaan [Roothaan 1951] and (independently) by Hall [Hall 1951]. The resulting equations are known as the Roothaan equations or the Roothaan–Hall equations. Unlike the integro-differential form of the Hartree–Fock equations, Equation (2.124), Roothaan and Hall recast the equations in matrix form, which can be solved using standard techniques and can be applied to systems of any geometry. We shall identify the major steps in the Roothaan approach,

starting with the expression for the Hartree–Fock energy for our closed-shell system, Equation (2.120):

$$E = 2\sum_{i=1}^{N/2} H_{ii}^{\text{core}} + \sum_{i=1}^{N/2}\sum_{j=1}^{N/2}(2J_{ij} - K_{ij}) \tag{2.146}$$

The corresponding Fock operator is (Equation (2.132)):

$$\mathscr{f}_i(1) = \mathscr{H}^{\text{core}}(1) + \sum_{j=1}^{N/2}\{2\mathscr{J}_j(1) - \mathscr{K}_j(1)\} \tag{2.147}$$

We now introduce the atomic orbital expansion for the orbitals ψ_i and substitute for the corresponding spin orbital χ_i into the Hartree–Fock equation, $\mathscr{f}_i(1)\chi_i(1) = \varepsilon_i\chi_i(1)$:

$$\mathscr{f}_i(1)\sum_{\nu=1}^{K} c_{\nu i}\phi_\nu(1) = \varepsilon_i\sum_{\nu=1}^{K} c_{\nu 1}\phi_\nu(1) \tag{2.148}$$

Pre-multiplying each side by $\phi_\mu(1)$ (where ϕ_μ is also a basis function) and integrating gives the following matrix equation:

$$\sum_{\nu=1}^{K} c_{\nu 1}\int d\nu_1\phi_\mu(1)\mathscr{f}_i(1)\phi_\nu(1) = \varepsilon_i\sum_{\nu=1}^{K} c_{\nu 1}\int d\nu_1\phi_\mu(1)\phi_\nu(1) \tag{2.149}$$

$\int d\nu_1\phi_\mu(1)\phi_\nu(1)$ is the overlap integral between the basis functions μ and ν, written $S_{\mu\nu}$. Unlike the molecular orbitals, which will be required to be orthonormal, the overlap between two basis functions is not necessarily zero (for example, they may be located on different atoms).

The elements of the *Fock matrix* are given by

$$F_{\mu\nu} = \int d\nu_1\phi_\mu(1)\mathscr{f}_i(1)\phi_\nu(1) \tag{2.150}$$

The Fock matrix elements for a closed-shell system can be expanded as follows by substituting the expression for the Fock operator:

$$F_{\mu\nu} = \int d\nu_1\phi_\mu(1)\mathscr{H}^{\text{core}}(1)\phi_\nu(1) + \sum_{j=1}^{N/2}\int d\nu_1\phi_\mu(1)[2\mathscr{J}_j(1) - \mathscr{K}_j(1)]\phi_\nu(1) \tag{2.151}$$

The elements of the Fock matrix can thus be written as the sum of core, Coulomb and exchange contributions. The core contribution is:

$$\int d\nu_1\phi_\mu(1)\mathscr{H}^{\text{core}}(1)\phi_\nu(1) = \int d\nu_1\phi_\mu(1)\left[-\frac{1}{2}\nabla^2 - \sum_{A=1}^{M}\frac{Z_A}{|r_1 - R_A|}\right]\phi_\nu(1) \equiv H_{\mu\nu}^{\text{core}} \tag{2.152}$$

The core contributions thus require the calculation of integrals that involve basis functions on up to two centres (depending upon whether ϕ_μ and ϕ_ν are centred on the same nucleus or not). Each element $H_{\mu\nu}^{\text{core}}$ can in turn be obtained as the sum of a kinetic energy integral and a potential energy integral corresponding to the two terms in the one-electron Hamiltonian.

The Coulomb and exchange contributions to the Fock matrix element $F_{\mu\nu}$ are together given by:

$$\sum_{j=1}^{N/2} \int d\nu_1 \phi_\mu(1)[2\mathscr{J}_j(1) - \mathscr{K}_j(1)]\phi_\nu(1) \tag{2.153}$$

Recall that the Coulomb operator $\mathscr{J}_j(1)$ due to interaction with a spin orbital χ_j is given by:

$$\mathscr{J}_j(1) = \int d\tau_2 \chi_j(2) \frac{1}{r_{12}} \chi_j(2) \tag{2.154}$$

We need to write each of the two occurrences of the spin orbital χ_j in this integral in terms of the appropriate linear combination of basis functions:

$$\mathscr{J}_j(1) = \int d\tau_2 \sum_{\sigma=1}^{K} c_{\sigma j}\phi_\sigma(2) \frac{1}{r_{12}} \sum_{\lambda=1}^{K} c_{\lambda j}\phi_\lambda(2) \tag{2.155}$$

We have used the indices σ and λ for the basis functions here. Similarly, the exchange contribution can be written:

$$\mathscr{K}_j(1)\chi_i(1) = \left[\int d\tau_2 \sum_{\sigma=1}^{K} c_{\sigma j}\phi_\sigma(2) \frac{1}{r_{12}} \chi_i(2)\right] \sum_{\lambda=1}^{K} c_{\lambda j}\phi_\lambda(2) \tag{2.156}$$

When the Coulomb and exchange operators are expressed in terms of the basis functions and the orbital expansion is substituted for χ_i, then their contributions to the Fock matrix element $F_{\mu\nu}$ take the following form:

$$\begin{aligned}
&\sum_{j=1}^{N/2} \int d\nu_1 \phi_\mu(1)[2\mathscr{J}_i(1) - \mathscr{K}_j(1)]\phi_\nu(1) \\
&\quad = \sum_{j=1}^{N/2} \sum_{\lambda=1}^{K} \sum_{\sigma=1}^{K} c_{\lambda j} c_{\sigma j} \left[\begin{array}{l} 2\int d\nu_1\, d\nu_2 \phi_\mu(1)\phi_\nu(1) \dfrac{1}{r_{12}} \phi_\lambda(2)\phi_\sigma(2) \\ -\int d\nu_1\, d\nu_2 \phi_\mu(1)\phi_\lambda(1) \dfrac{1}{r_{12}} \phi_\nu(2)\phi_\sigma(2) \end{array} \right] \\
&\quad \equiv \sum_{j=1}^{N/2} \sum_{\lambda=1}^{K} \sum_{\sigma=1}^{K} c_{\lambda j} c_{\sigma j} [2(\mu\nu|\lambda\sigma) - (\mu\lambda|\nu\sigma)]
\end{aligned} \tag{2.157}$$

We have used the shorthand notation for the integrals in the final expression. Note that the two-electron integrals may involve up to four different basis functions (μ, ν, λ, σ), which may in turn be located at four different centres. This has important consequences for the way in which we try to solve the equations.

It is helpful to simplify Equation (2.157) by introducing the *charge density matrix*, **P**, whose elements are defined as:

$$P_{\mu\nu} = 2\sum_{i=1}^{N/2} c_{\mu i} c_{\nu i} \quad \text{and} \quad P_{\lambda\sigma} = 2\sum_{i=1}^{N/2} c_{\lambda i} c_{\sigma i} \tag{2.158}$$

Note that the summations are over the $N/2$ occupied orbitals. Other properties can be calculated from the density matrix; for example, the electronic energy is:

$$E = \frac{1}{2}\sum_{\mu=1}^{K}\sum_{\nu=1}^{K} P_{\mu\nu}(H_{\mu\nu}^{\text{core}} + F_{\mu\nu}) \tag{2.159}$$

The electron density at a point $\mathbf{r}$ can also be expressed in terms of the density matrix:

$$\rho(\mathbf{r}) = \sum_{\mu=1}^{K}\sum_{\nu=1}^{K} P_{\mu\nu}\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r}) \tag{2.160}$$

The expression for each element $F_{\mu\nu}$ of the Fock matrix elements for a closed-shell system of N electrons then becomes:

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + \sum_{\lambda=1}^{K}\sum_{\sigma=1}^{K} P_{\lambda\sigma}[(\mu\nu|\lambda\sigma) - \tfrac{1}{2}(\mu\lambda|\nu\sigma)] \tag{2.161}$$

This is the standard form for the expression for the Fock matrix in the Roothaan–Hall equations

2.5.4 Solving the Roothaan–Hall Equations

The Fock matrix is a $K \times K$ square matrix that is symmetric if real basis functions are used. The Roothaan–Hall equations (2.149) can be conveniently written as a matrix equation:

$$\mathbf{FC} = \mathbf{SCE} \tag{2.162}$$

The elements of the $K \times K$ matrix $\mathbf{C}$ are the coefficients $c_{\nu i}$:

$$\mathbf{C} = \begin{pmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,K} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,K} \\ \vdots & \vdots & & \vdots \\ c_{K,1} & c_{K,2} & \cdots & c_{K,K} \end{pmatrix} \tag{2.163}$$

$\mathbf{E}$ is a diagonal matrix whose elements are the orbital energies:

$$\mathbf{E} = \begin{pmatrix} \varepsilon_1 & 0 & \cdots & 0 \\ 0 & \varepsilon_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varepsilon_K \end{pmatrix} \tag{2.164}$$

Let us consider how we might solve the Roothaan–Hall equations and thereby obtain the molecular orbitals. The first point we must note is that the elements of the Fock matrix, which appear on the left-hand side of Equation (2.162), depend on the molecular orbital coefficients $c_{\nu i}$, which also appear on the right-hand side of the equation. Thus an iterative procedure is required to find a solution.

The one-electron contributions $H_{\mu\nu}^{\text{core}}$ due to the electrons moving in the field of the bare nuclei do not depend on the basis set coefficients and remain unchanged throughout the calculation. However, the Coulomb and exchange contributions do depend on the coefficients and we would expect these to vary throughout the calculation. The individual two-electron integrals $(\mu\nu|\lambda\sigma)$ are, however, constant throughout the calculation. An obvious strategy is thus to calculate and store these integrals for later use.

Having written the Roothaan–Hall equations in matrix form we would obviously like to solve them using standard matrix eigenvalue methods (discussed in Section 1.10.3). However, standard eigenvalue methods would require an equation of the form $\mathbf{FC} = \mathbf{CE}$. The Roothaan–Hall equations only adopt such a form if the overlap matrix, $\mathbf{S}$, is equal to the unit matrix, $\mathbf{I}$ (in which all diagonal elements are equal to 1 and all off-diagonal elements are zero). The functions ϕ are usually normalised but they are not necessarily orthogonal (for example, because they are located on different atoms) and so there will invariably be non-zero off-diagonal elements of the overlap matrix. To solve the Roothaan–Hall equations using standard methods they must be transformed. This corresponds to transforming the basis functions so that they form an orthonormal set. We seek a matrix $\mathbf{X}$, such that $\mathbf{X}^{\mathrm{T}}\mathbf{SX} = \mathbf{I}$. $\mathbf{X}^{\mathrm{T}}$ is the transpose of $\mathbf{X}$, obtained by interchanging rows and columns. There are various ways in which $\mathbf{X}$ can be calculated; in *symmetric orthogonalisation*, the overlap matrix is diagonalised. Diagonalisation involves finding the matrix $\mathbf{U}$ such that

$$\mathbf{U}^{\mathrm{T}}\mathbf{SU} = \mathbf{D} = \mathrm{diag}(\lambda_1 \ldots \lambda_K) \tag{2.165}$$

$\mathbf{D}$ is the diagonal matrix containing the eigenvalues λ_i of $\mathbf{S}$, and $\mathbf{U}$ contains the eigenvectors of $\mathbf{S}$. $\mathbf{U}^{\mathrm{T}}$ is the transpose of the matrix $\mathbf{U}$. (This expression is often written $\mathbf{U}^{-1}\mathbf{SU} = \mathbf{D}$ since for real basis functions $\mathbf{U}^{-1} = \mathbf{U}^{\mathrm{T}}$.) Then the matrix $\mathbf{X}$ is given by $\mathbf{X} = \mathbf{U}\mathbf{D}^{-1/2}\mathbf{U}^{\mathrm{T}}$, where $\mathbf{D}^{-1/2}$ is formed from the inverse square roots of $\mathbf{D}$. We shall write $\mathbf{X}$ as $\mathbf{S}^{-1/2}$, as it can be considered to be the inverse square root of the overlap matrix: $\mathbf{S}^{-1/2}\mathbf{S}\mathbf{S}^{-1/2} = \mathbf{I}$.

The Roothaan–Hall equations can now be manipulated as follows. Both sides of Equation (2.162) are pre-multiplied by the matrix $\mathbf{S}^{-1/2}$:

$$\mathbf{S}^{-1/2}\mathbf{FC} = \mathbf{S}^{-1/2}\mathbf{SCE} = \mathbf{S}^{1/2}\mathbf{CE} \tag{2.166}$$

Inserting the unit matrix, in the form $\mathbf{S}^{-1/2}\mathbf{S}^{1/2}$, into the left-hand side gives:

$$\mathbf{S}^{-1/2}\mathbf{F}(\mathbf{S}^{-1/2}\mathbf{S}^{1/2})\mathbf{C} = \mathbf{S}^{1/2}\mathbf{CE} \tag{2.167}$$

or

$$\mathbf{S}^{-1/2}\mathbf{F}\mathbf{S}^{-1/2}(\mathbf{S}^{1/2}\mathbf{C}) = (\mathbf{S}^{1/2}\mathbf{C})\mathbf{E} \tag{2.168}$$

Equation (2.168) can be written $\mathbf{F}'\mathbf{C}' = \mathbf{C}'\mathbf{E}$, where $\mathbf{F}' = \mathbf{S}^{-1/2}\mathbf{F}\mathbf{S}^{-1/2}$ and $\mathbf{C}' = \mathbf{S}^{1/2}\mathbf{C}$.

The matrix equation $\mathbf{F}'\mathbf{C}' = \mathbf{C}'\mathbf{E}$ can be solved using standard methods; a solution only exists if the determinant $|\mathbf{F}' - \mathbf{EI}|$ equals zero. In simple cases this can be done by multiplying out the determinant to give a polynomial (the secular equation) whose roots are the eigenvalues ε_i, but for large matrices a much more practical approach involves the diagonalisation of $\mathbf{F}'$. The matrix of coefficients, $\mathbf{C}'$, are the eigenvectors of $\mathbf{F}'$. The basis function coefficients $\mathbf{C}$ can then be obtained from $\mathbf{C}'$ using $\mathbf{C} = \mathbf{S}^{-1/2}\mathbf{C}'$. A common scheme for

solving the Roothaan-Hall equations is thus as follows:

1. Calculate the integrals to form the Fock matrix, **F**.
2. Calculate the overlap matrix, **S**.
3. Diagonalise **S**.
4. Form $\mathbf{S}^{-1/2}$.
5. Guess, or otherwise calculate, an initial density matrix, **P**.
6. Form the Fock matrix using the integrals and the density matrix **P**.
7. Form $\mathbf{F}' = \mathbf{S}^{-1/2}\mathbf{F}\mathbf{S}^{-1/2}$.
8. Solve the secular equation $|\mathbf{F}' - \mathbf{E}\mathbf{I}| = 0$ to give the eigenvalues **E** and the eigenvectors $\mathbf{C}'$ by diagonalising $\mathbf{F}'$.
9. Calculate the molecular orbital coefficients, **C** from $\mathbf{C} = \mathbf{S}^{-1/2}\mathbf{C}'$.
10. Calculate a new density matrix, **P**, from the matrix **C**.
11. Check for convergence. If the calculation has converged, stop. Otherwise repeat from step 6 using the new density matrix, **P**.

This procedure requires an initial guess of the density matrix, **P**. The simplest approach is to use the null matrix, which corresponds to ignoring all the electron–electron terms so that the electrons just experience the bare nuclei. This can sometimes lead to convergence problems, which may be prevented if a lower level of theory (such as semi-empirical or extended Hückel) is used to provide the initial guess. Moreover, a better guess may enable the calculation to be performed more quickly. A variety of criteria can be used to establish whether the calculation has converged or not. For example, the density matrix can be compared with that from the previous iteration, and/or the change in energy can be monitored together with the basis set coefficients.

The result of a Hartree-Fock calculation is a set of K molecular orbitals, where K is the number of basis functions in the calculation. The N electrons are then fed into these orbitals in accordance with the Aufbau principle, two electrons per orbital, starting with the lowest-energy orbitals. The remaining orbitals do not contain any electrons; these are known as the *virtual orbitals*. Alternative electronic configurations can be generated by exciting electrons from the occupied orbitals to the virtual orbitals; these excited configurations are used in more advanced calculations that will be discussed in Chapter 3.

A Hartree–Fock calculation provides a set of orbital energies, ε_i. What is the significance of these? The energy of an electron in a spin orbital is calculated by adding the core interaction H_{ii}^{core} to the Coulomb and exchange interactions with the other electrons in the system:

$$\varepsilon_i = H_{ii}^{\text{core}} + \sum_{j=1}^{N/2}(2J_{ij} - K_{ij}) \tag{2.169}$$

The total electronic energy of the ground state is given by Equation (2.120):

$$E = 2\sum_{i=1}^{N/2} H_{ii}^{\text{core}} + \sum_{i=1}^{N/2}\sum_{j=1}^{N/2}(2J_{ij} - K_{ij}) \tag{2.170}$$

The total energy is therefore not equal to the sum of the individual orbital energies but is related as follows:

$$E = \sum_{i=1}^{N} \varepsilon_i - \sum_{i=1}^{N/2} \sum_{j=1}^{N/2} (2J_{ij} - K_{ij}) \tag{2.171}$$

The reason for the discrepancy is that the individual orbital energies include contributions from the interaction between that electron and all the nuclei and all other electrons in the system. The Coulomb and exchange interactions between pairs of electrons are therefore counted twice when summing the individual orbital energies.

2.5.5 A Simple Illustration of the Roothaan–Hall Approach

We will illustrate the stages involved in the Roothaan–Hall approach using the helium hydrogen molecular ion, HeH^+, as an example. This is a two-electron system. Our objective here is to show how the Roothaan–Hall method can be used to derive the wavefunction, for a fixed internuclear distance of 1 Å. We use HeH^+ rather than H_2 as our system as the lack of symmetry in HeH^+ makes the procedure more informative. There are two basis functions, $1s_A$ (centred on the helium atom) and $1s_B$ (on the hydrogen). The numerical values of the integrals that we shall use in our calculation were obtained using a Gaussian series approximation to the Slater orbitals (the STO-3G basis set, which is described in Section 2.6). This detail need not concern us here. Each wavefunction is expressed as a linear combination of the two 1s atomic orbitals centred on the nuclei A and B:

$$\psi_1 = c_{1A}1s_A + c_{1B}1s_B \tag{2.172}$$

$$\psi_2 = c_{2A}1s_A + c_{2B}1s_B \tag{2.173}$$

First, it is necessary to calculate the various one- and two-electron integrals and to formulate the Fock and overlap matrices, each of which will be a 2×2 symmetric matrix (as there are two orbitals in the basis set). The diagonal elements of the overlap matrix, **S**, are equal to 1.0 as each basis function is normalised; the off-diagonal elements have smaller, but non-zero, values that are equal to the overlap between $1s_A$ and $1s_B$ for the internuclear distance chosen. The matrix **S** is:

$$\mathbf{S} = \begin{pmatrix} 1.0 & 0.392 \\ 0.392 & 1.0 \end{pmatrix} \tag{2.174}$$

The core contributions $H_{\mu\nu}^{\text{core}}$ can be calculated as the sum of three 2×2 matrices comprising the kinetic energy (T) and nuclear attraction terms for the two nuclei A and B ($\mathbf{V}_A$ and $\mathbf{V}_B$). The elements of these three matrices are obtained by evaluating the following integrals:

$$\mathbf{T}_{\mu\nu} = \int d\nu_1 \phi_\mu(1)(-\tfrac{1}{2}\nabla^2)\phi_\nu(1)$$

$$\mathbf{V}_{A,\mu\nu} = \int d\nu_1 \phi_\mu(1)\left(-\frac{Z_A}{r_{1A}}\right)\phi_\nu(1)$$

$$\mathbf{V}_{B,\mu\nu} = \int d\nu_1 \phi_\mu(1)\left(-\frac{Z_B}{r_{1B}}\right)\phi_\nu(1) \tag{2.175}$$

The matrices are:

$$\mathbf{T} = \begin{pmatrix} 1.412 & 0.081 \\ 0\,081 & 0.760 \end{pmatrix} \quad \mathbf{V}_A = \begin{pmatrix} -3\,344 & -0.758 \\ -0.758 & -1.026 \end{pmatrix} \quad \mathbf{V}_B = \begin{pmatrix} -0.525 & -0.308 \\ -0.308 & -1.227 \end{pmatrix} \tag{2.176}$$

$\mathbf{H}^{\text{core}}$ is the sum of these three:

$$\mathbf{H}^{\text{core}} = \begin{pmatrix} -2.457 & -0.985 \\ -0.985 & -1.493 \end{pmatrix} \tag{2.177}$$

As far as the two-electron integrals are concerned, with two basis functions there are a total of 16 possible two-electron integrals. There are however only six unique two-electron integrals, as the indices can be permuted as follows:

(i) $(1s_A 1s_A|1s_A 1s_A) = 1.056$
(ii) $(1s_A 1s_A|1s_A 1s_B) = (1s_A 1s_A|1s_B 1s_A) = (1s_A 1s_B|1s_A 1s_A)$
$= (1s_B 1s_A|1s_A 1s_A) = 0.303$
(iii) $(1s_A 1s_B|1s_A 1s_B) = (1s_A 1s_B|1s_B 1s_A) = (1s_B 1s_A|1s_A 1s_B)$
$= (1s_B 1s_A|1s_B 1s_A) = 0.112$
(iv) $(1s_A 1s_A|1s_B 1s_B) = (1s_B 1s_B|1s_A 1s_A) = 0.496$
(v) $(1s_A 1s_B|1s_B 1s_B) = (1s_B 1s_A|1s_B 1s_B) = (1s_B 1s_B|1s_A 1s_B)$
$= (1s_B 1s_B|1s_B 1s_A) = 0.244$
(vi) $(1s_B 1s_B|1s_B 1s_B) = 0.775$

To reiterate, these integrals are calculated as follows:

$$(\mu\nu|\lambda\sigma) = \iint d\nu_1\, d\nu_2 \phi_\mu(1)\phi_\nu(1)\frac{1}{r_{12}}\phi_\lambda(2)\phi_\sigma(2) \tag{2.178}$$

Having calculated the integrals, we are now ready to start the SCF calculation. To formulate the Fock matrix it is necessary to have an initial guess of the density matrix, **P**. The simplest approach is to use the null matrix in which all elements are zero. In this initial step the Fock matrix **F** is therefore equal to $\mathbf{H}^{\text{core}}$.

The Fock matrix must next be transformed to $\mathbf{F}'$ by pre- and post-multiplying by $\mathbf{S}^{-1/2}$:

$$\mathbf{S}^{-1/2} = \begin{pmatrix} -1.065 & -0.217 \\ -0.217 & 1.065 \end{pmatrix} \tag{2 179}$$

$\mathbf{F}'$ for this first iteration is thus:

$$\mathbf{F}' = \begin{pmatrix} -2.401 & -0.249 \\ -0.249 & -1.353 \end{pmatrix} \tag{2.180}$$

Diagonalisation of $\mathbf{F}'$ gives its eigenvalues and eigenvectors, which are:

$$\mathbf{E} = \begin{pmatrix} -2.458 & 0.0 \\ 0\,0 & -1.292 \end{pmatrix} \quad \mathbf{C}' = \begin{pmatrix} 0\,975 & -0.220 \\ 0.220 & 0\,975 \end{pmatrix} \tag{2.181}$$

The coefficients **C** are obtained from $\mathbf{C} = \mathbf{S}^{-1/2}\mathbf{C}'$ and are thus:

$$\mathbf{C} = \begin{pmatrix} 0.991 & -0.446 \\ 0.022 & 1.087 \end{pmatrix} \tag{2.182}$$

To formulate the density matrix **P** we need to identify the occupied orbital(s). With a two-electron system both electrons occupy the orbital with the lowest energy (i.e. the orbital with the lowest eigenvalue). At this stage the lowest-energy orbital is:

$$\psi = 0.991\ 1s_A + 0.022\ 1s_B \tag{2.183}$$

The orbital is composed largely of the s orbital on the helium nucleus; in the absence of any electron–electron repulsion the electrons tend to congregate near the nucleus with the larger charge. The density matrix corresponding to this initial wavefunction is:

$$\mathbf{P} = \begin{pmatrix} 1.964 & 0.044 \\ 0.044 & 0.001 \end{pmatrix} \tag{2 184}$$

The new Fock matrix is formed using **P** and the two-electron integrals together with $\mathbf{H}^{core}$. For example, the element F_{11} is given by:

$$\begin{aligned} F_{11} = H_{11}^{core} &+ P_{11}[(1s_A1s_A|1s_A1s_A) - \tfrac{1}{2}(1s_A1s_A|1s_A1s_A)] \\ &+ P_{12}[(1s_A1s_A|1s_A1s_B) - \tfrac{1}{2}(1s_A1s_A|1s_A1s_B)] \\ &+ P_{21}[(1s_A1s_A|1s_B1s_B) - \tfrac{1}{2}(1s_A1s_B|1s_A1s_B)] \\ &+ P_{12}[(1s_A1s_A|1s_B1s_B) - \tfrac{1}{2}(1s_A1s_B|1s_A1s_B)] \end{aligned} \tag{2.185}$$

The complete Fock matrix is:

$$\mathbf{F} = \begin{pmatrix} -1.406 & -0.690 \\ -0\ 690 & -0.618 \end{pmatrix} \tag{2 186}$$

The energy that corresponds to this Fock matrix (calculated using Equation (2.159)) is −3 870 Hartree. In the next iteration, the various matrices are as follows:

$$\mathbf{F}' = \begin{pmatrix} -1.305 & -0.347 \\ -0.347 & -0.448 \end{pmatrix} \quad \mathbf{E} = \begin{pmatrix} -1.427 & 0.0 \\ 0.0 & -0.325 \end{pmatrix}$$

$$\mathbf{C}' = \begin{pmatrix} 0.943 & -0.334 \\ 0.334 & 0.943 \end{pmatrix} \quad \mathbf{C} = \begin{pmatrix} 0.931 & -0.560 \\ 0.150 & 1.076 \end{pmatrix}$$

$$\mathbf{P} = \begin{pmatrix} 1\,735 & 0.280 \\ 0.280 & 0\,045 \end{pmatrix} \quad \mathbf{F} = \begin{pmatrix} -1.436 & -0.738 \\ -0.738 & -0.644 \end{pmatrix}$$

$$\text{Energy} = -3.909 \text{ Hartree} \tag{2.187}$$

The calculation proceeds as illustrated in Table 2.2, which shows the variation in the coefficients of the atomic orbitals in the lowest-energy wavefunction and the energy for the first four SCF iterations. The energy is converged to six decimal places after six iterations and the charge density matrix after nine iterations.

The final wavefunction still contains a large proportion of the 1s orbital on the helium atom, but less than was obtained without the two-electron integrals.

Iteration	$c(1s_A)$	$c(1s_B)$	Energy
1	0 991	0.022	−3.870
2	0.931	0.150	−3.909
3	0 915	0.181	−3.911
4	0.912	0.187	−3.911

Table 2 2 Variation in basis set coefficients and electronic energy for the HeH^+ molecule.

2.5.6 Application of the Hartree–Fock Equations to Molecular Systems

We are now in a position to consider how the Hartree–Fock theory we have developed can be used to perform practical quantum mechanical calculations on molecular systems. This is an appropriate place in our discussion to distinguish the two major categories of quantum mechanical molecular orbital calculations: the *ab initio* and the semi-empirical methods. *Ab initio* strictly means 'from the beginning', or 'from first principles', which would imply that a calculation using such an approach would require as input only physical constants such as the speed of light, Planck's constant, the masses of elementary particles, and so on. *Ab initio* in fact usually refers to a calculation which uses the full Hartree–Fock/Roothaan–Hall equations, without ignoring or approximating any of the integrals or any of the terms in the Hamiltonian. The *ab initio* methods do rely upon calibration calculations, and this has led some quantum chemists, notably Dewar (who has played a large part in the development of semi-empirical methods), to claim that any real difference between the *ab initio* and semi-empirical methods is entirely pedagogical. By contrast, semi-empirical methods simplify the calculations, using parameters for some of the integrals and/or ignoring some of the terms in the Hamiltonian. First we shall consider *ab initio* methods.

2.6 Basis Sets

The basis sets most commonly used in quantum mechanical calculations are composed of atomic functions. An obvious choice would be the Slater type orbitals. Unfortunately, Slater functions are not particularly amenable to implementation in molecular orbital calculations. This is because some of the integrals are difficult, if not impossible, to evaluate, particularly when the atomic orbitals are centred on different nuclei. It is relatively straightforward to calculate integrals involving one or two centres, such as $(\mu\mu|\nu\nu)$, $(\mu\nu|\nu\nu)$ and $(\mu\nu|\mu\nu)$. Three- and four-centre integrals are also feasible with Slater functions if the atomic orbitals are located on the same atom. However, three- and four-centre integrals are very difficult if the atomic orbitals are based on different atoms. It is common in *ab initio* calculations to replace the Slater orbitals by functions based upon Gaussians. A Gaussian function has the form $\exp(-\alpha r^2)$, and *ab initio* calculations use basis functions comprising integral powers of x, y and z multiplied by $\exp(-\alpha r^2)$:

$$x^a y^b z^c \exp(-\alpha r^2) \tag{2.188}$$

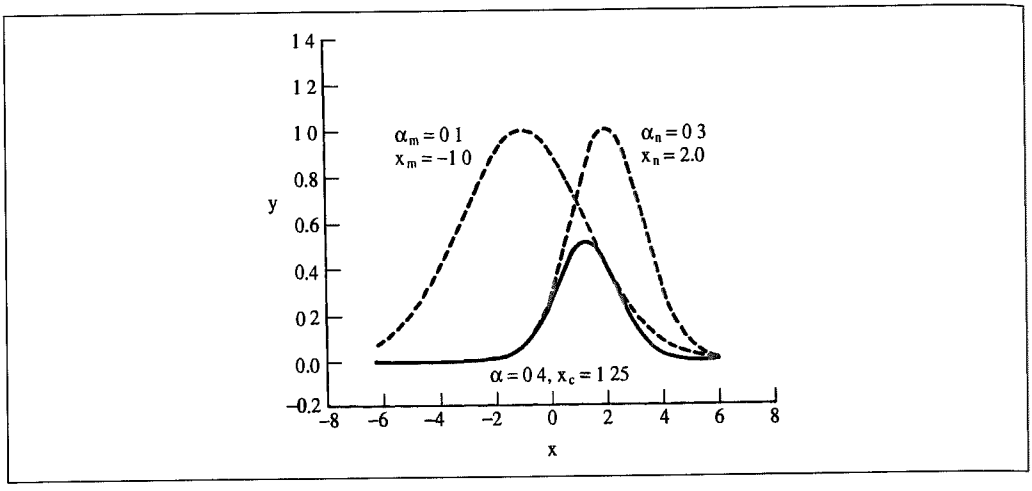

Fig 2 4: The product of two Gaussian functions is another Gaussian centred along the line joining their centres In this case the equations of the two functions are $y = exp[-0\,1(x + 1\,0)^2]$ *and* $y = exp[-0\,3(x - 2.0)^2]$ *and the equation of the product is* $y = exp(-27/40)[-0\,4(x - 1\,25)^3]$ *(Equation (2.189))*

α determines the radial extent (or 'spread') of a Gaussian function; a function with a large value of α does not spread very far, whereas a small value of α gives a large spread. The *order* of these Gaussian-type functions is determined by the powers of the Cartesian variables; a zeroth-order function has $a + b + c = 0$; a first-order function has $a + b + c = 1$, and so on. There is thus one zeroth-order function, three first-order functions and six second-order functions. The idea of using Gaussian functions in quantum mechanical calculations is often ascribed to Boys [Boys 1950]. A major advantage of Gaussian functions is that the product of two Gaussians can be expressed as a single Gaussian, located along the line joining the centres of the two Gaussians m and n (Figure 2.4):

$$\exp(-\alpha_m r_m^2)\exp(-\alpha_n r_n^2) = \exp\left(-\frac{\alpha_m\alpha_n}{\alpha_m + \alpha_n} r_{mn}^2\right)\exp(-\alpha r_c^2) \tag{2.189}$$

r_{mn} is the distance between the centres m and n, and the orbital exponent α of the combined function is related to the exponents α_m and α_n by:

$$\alpha = \alpha_m + \alpha_n \tag{2.190}$$

r_C is the distance from point C, which has coordinates:

$$x_c = \frac{\alpha_m x_m + \alpha_n x_n}{\alpha_m + \alpha_n}; \quad y_c = \frac{\alpha_m y_m + \alpha_n y_n}{\alpha_m + \alpha_n}; \quad z_c = \frac{\alpha_m z_m + \alpha_n z_n}{\alpha_m + \alpha_n} \tag{2.191}$$

x_m, y_m, z_m and x_n, y_n, z_n are the centres of the two original Gaussians m and n respectively.

Thus, in a two-electron integral of the form $(\mu\nu|\lambda\sigma)$, the product $\phi_\mu(1)\phi_\nu(1)$ (where ϕ_μ and ϕ_ν may be on different centres) can be replaced by a single Gaussian function that is centred at the appropriate point C. For Cartesian Gaussian functions the calculation is more complicated than for the example we have stated above, due to the presence of the Cartesian functions, but even so, efficient methods for performing the integrals have been devised.

The zeroth-order Gaussian function g_s has s-orbital angular symmetry; the three first-order Gaussian functions have p-orbital symmetry. In normalised form these are:

$$g_s(\alpha, r) = \left(\frac{2\alpha}{\pi}\right)^{3/4} e^{-\alpha r^2} \tag{2.192}$$

$$g_x(\alpha, r) = \left(\frac{128\alpha^5}{\pi^3}\right)^{1/4} x\, e^{-\alpha r^2} \tag{2.193}$$

$$g_y(\alpha, r) = \left(\frac{128\alpha^5}{\pi^3}\right)^{1/4} y\, e^{-\alpha r^2} \tag{2.194}$$

$$g_z(\alpha, r) = \left(\frac{128\alpha^5}{\pi^3}\right)^{1/4} z\, e^{-\alpha r^2} \tag{2.195}$$

The six second-order functions have the following form, exemplified by two of the functions:

$$g_{xx}(\alpha, r) = \left(\frac{2048\alpha^7}{9\pi^3}\right)^{1/4} x^2\, e^{-\alpha r^2} \tag{2.196}$$

$$g_{xy}(\alpha, r) = \left(\frac{2048\alpha^7}{9\pi^3}\right)^{1/4} xy\, e^{-\alpha r^2} \tag{2.197}$$

These second-order functions do not all have the same angular symmetry as the 3d atomic orbitals, but a set comprising g_{xy}, g_{xz} and g_{yz}, together with two linear combinations of the g_{xx}, g_{yy} and g_{zz}, does give the desired result:

$$g_{3zz-rr} = \tfrac{1}{2}(2g_{zz} - g_{xx} - g_{yy}) \tag{2.198}$$

$$g_{xx-yy} = \sqrt{\tfrac{3}{4}}(g_{xx} - g_{yy}) \tag{2.199}$$

The remaining sixth linear combination has the symmetry properties of an s function:

$$g_{rr} = \sqrt{5}(g_{xx} + g_{yy} + g_{zz}) \tag{2.200}$$

The advantages of Gaussian functions are countered by some serious shortcomings. This can be readily seen from a graphical comparison of the 1s Slater function and its 'best' Gaussian approximation, Figure 2.5. Unlike the Slater functions the Gaussian functions do not have a cusp at the origin and they also decay towards zero more quickly. It is found that replacing a Slater type orbital by a single Gaussian function leads to unacceptable errors. However, this problem can be overcome if each atomic orbital is represented as a linear combination of Gaussian functions. Each linear combination has the following form:

$$\phi_\mu = \sum_{i=1}^{L} d_{i\mu}\phi_i(\alpha_{i\mu}) \tag{2.201}$$

$d_{i\mu}$ is the coefficient of the primitive Gaussian function ϕ_i, which has exponent $\alpha_{i\mu}$. L is the number of functions in the expansion. For example, the linear combinations of Gaussian 1s functions that can be used to represent a 1s Slater type orbital with exponent $\xi = 1$ are given in Table 2.3.

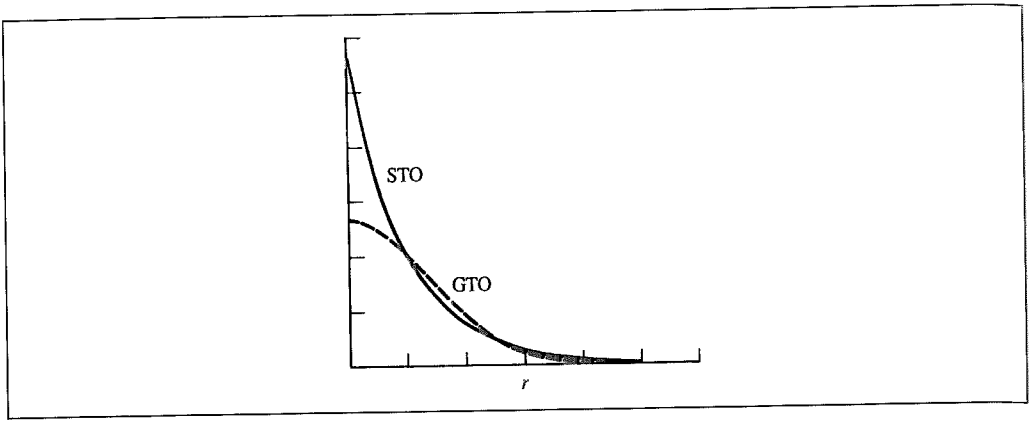

Fig. 2 5· The 1s Slater type orbital and the best Gaussian equivalent.

The coefficients and the exponents are found by least-squares fitting, in which the overlap between the Slater type function and the Gaussian expansion is maximised. Thus, for the 1s Slater type orbital we seek to maximise the following integral:

$$S = \frac{1}{\sqrt{\pi}} \left(\frac{2\alpha}{\pi} \right)^{3/4} \int dr\, e^{-r}\, e^{-\alpha r^2} \tag{2.202}$$

A graphical comparison of the 1s Slater type orbital and the four Gaussian expansions in Table 2.3 is shown in Figure 2.6. It is clear that the fit improves as the number of Gaussian functions increases, but even so, the addition of many more Gaussian functions cannot properly describe the exponential tail in the 'true' function and the cusp at the nucleus. This means that Gaussian functions underestimate the long-range overlap between atoms and the charge and spin density at the nucleus.

A Gaussian expansion contains two parameters: the coefficient and the exponent. The most flexible way to use Gaussian functions in *ab initio* molecular orbital calculations permits both of these parameters to vary during the calculation. Such a calculation is said to use

Number of Gaussians	Exponent, α	Expansion coefficient, *d*
1	0.270 950	1.00
2	0.151 623	0 678 914
	0.851 819	0.430 129
3	0.109 818	0.444 635
	0.405 771	0.535 28
	2 227 66	0.154 329
4	0.088 0187	0 291 626
	0 265 204	0.532 846
	0 954 620	0 260 141
	5.216 86	0.056 7523

Table 2 3 Coefficients and exponents for best-fit Gaussian expansions for the 1s Slater type orbital [Hehre et al *1969]*

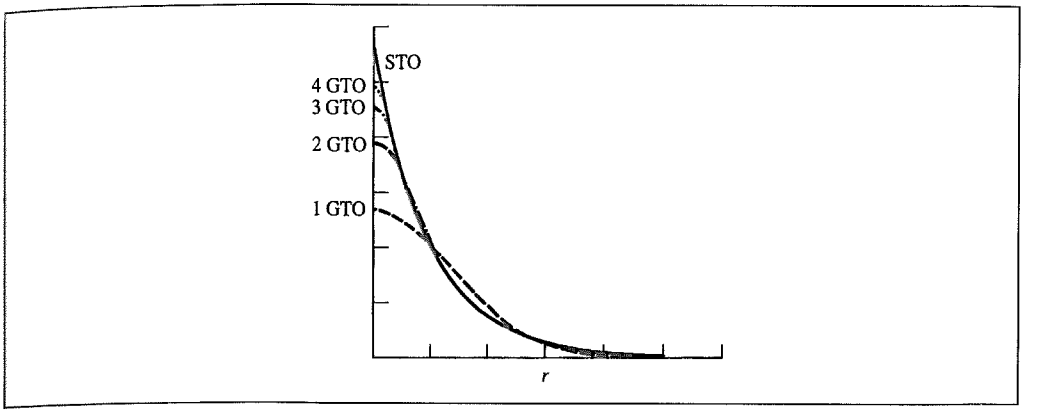

Fig 2 6: Comparison of 1s Slater type orbital and Gaussian expansions with up to four terms

uncontracted or *primitive* Gaussians. However, calculations with primitive Gaussians require a significant computational effort and so basis sets that consist of *contracted* Gaussian functions are most commonly employed. In a contracted function the contraction coefficients and exponents are pre-determined and remain constant during the calculation. The series of Gaussian functions in such cases is commonly referred to as a *contraction*, with the *contraction length* being the number of terms in the expansion. A further approximation that is often employed for the sake of computational efficiency is to use the same Gaussian exponents for the s and p orbitals in a given shell. This clearly restricts the flexibility of the basis set, but it does have the advantage of significantly reducing the number of numerically different integrals that need to be calculated.

Quantum chemists have devised efficient short-hand notation schemes to denote the basis set used in an *ab initio* calculation, although this does mean that a proliferation of abbreviations and acronyms are introduced. However, the codes are usually quite simple to understand. We shall concentrate on the notation used by Pople and co-workers in their Gaussian series of programs (see also the appendix to this chapter).

A *minimal basis set* is a representation that, strictly speaking, contains just the number of functions that are required to accommodate all the filled orbitals in each atom. In practice, a minimal basis set normally includes all of the atomic orbitals in the shell. Thus, for hydrogen and helium a single s-type function would be required; for the elements from lithium to neon the 1s, 2s and 2p functions are used, and so on. The basis sets STO-3G, STO-4G, etc. (in general, STO-*n*G) are all minimal basis sets in which *n* Gaussian functions are used to represent each orbital. It is found that at least three Gaussian functions are required to properly represent each Slater type orbital and so the STO-3G basis set is the 'absolute minimum' that should be used in an *ab initio* molecular orbital calculation. In fact, there is often little difference between the results obtained with the STO-3G basis set and the larger minimal basis sets with more Gaussian functions, although for hydrogen-bonded complexes STO-4G can perform significantly better. The STO-3G basis set does perform remarkably well in predicting molecular geometries, though this is due in part to

a fortuitous cancellation of errors. Of course, the computational effort increases with the number of functions in the Gaussian expansion.

The minimal basis sets are well known to have several deficiencies. There are particular problems with compounds containing atoms at the end of a period, such as oxygen or fluorine. Such atoms are described using the same number of basis functions as the atoms at the beginning of the period, despite the fact that they have more electrons. A minimal basis set only contains one contraction per atomic orbital and as the radial exponents are not allowed to vary during the calculation the functions cannot expand or contract in size in accordance with the molecular environment. The third drawback is that a minimal basis set cannot describe non-spherical aspects of the electronic distribution. For example, for a second-row element such as carbon the only functions that incorporate any anisotropy are the $2p_x$, $2p_y$ and $2p_z$ functions. As the radial components of these functions are required to be the same, no one component (x, y or z) can differ from another.

These problems with minimal basis sets can be addressed if more than one function is used for each orbital. A basis set which doubles the number of functions in the minimal basis set is described as a *double zeta* basis. Thus, a linear combination of a 'contracted' function and a 'diffuse' function gives an overall result that is intermediate between the two. The basis set coefficients of the contracted and the diffuse functions are automatically calculated by the SCF procedure, which thus automatically determines whether a more contracted or a more diffuse representation of that particular orbital is required. Such an approach can provide a solution to the anisotropy problem because it is then possible to have different linear combinations for the p_x, p_y and p_z orbitals.

An alternative to the double zeta basis approach is to double the number of functions used to describe the valence electrons but to keep a single function for the inner shells. The rationale for this approach is that the core orbitals, unlike the valence orbitals, do not affect chemical properties very much and vary only slightly from one molecule to another. The notation used for such *split valence* double zeta basis sets is exemplified by 3-21G. In this basis set three Gaussian functions are used to describe the core orbitals. The valence electrons are also represented by three Gaussians: the contracted part by two Gaussians and the diffuse part by one Gaussian. The most commonly used split valence basis sets are 3-21G, 4-31G and 6-31G.

Simply increasing the number of basis functions (triple zeta, quadruple zeta, etc.) does not necessarily improve the model. In fact, it can give rise to wholly erroneous results, particularly for molecules with a strongly anisotropic charge distribution. All of the basis sets we have encountered so far use functions that are centred on atomic nuclei. The use of split valence basis sets can help to surmount the problems with non-isotropic charge distribution but not completely. The charge distribution about an atom in a molecule is usually perturbed in comparison with the isolated atom. For example, the electron cloud in an isolated hydrogen atom is symmetrical, but when the hydrogen atom is present in a molecule the electrons are attracted towards the other nuclei. The distortion can be considered to correspond to mixing p-type character into the 1s orbital of the isolated atom to give a form of sp hybrid. In a similar manner, the unoccupied d orbitals introduce asymmetry into p orbitals (Figure 2.7). The most common solution to this problem is to introduce

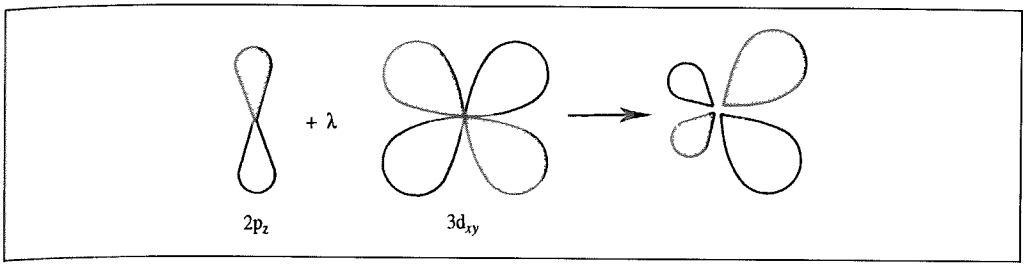

Fig. 2 7. The addition of a $3d_{xy}$ orbital to $2p_z$ gives a distorted orbital. (Figure adapted from Hehre W J, L Radom, P v R Schleyer and J A Hehre 1986 Ab initio Molecular Orbital Theory. *New York, Wiley)*

polarisation functions into the basis set. The polarisation functions have a higher angular quantum number and so correspond to p orbitals for hydrogen and d orbitals for the first- and second-row elements.

The use of polarisation basis functions is indicated by an asterisk (*). Thus, 6-31G* refers to a 6-31G basis set with polarisation functions on the heavy (i.e. non-hydrogen) atoms. Two asterisks (e.g. 6-31G**) indicate the use of polarisation (i.e. p) functions on hydrogen and helium. The 6-31G** basis set is particularly useful where hydrogen acts as a bridging atom. Partial polarisation basis sets have also been developed. For example, the 3-21G(*) basis set has the same set of Gaussians as the 3-21G basis set (i.e. three functions for the inner shell, two contracted functions and one diffuse function for the valence shell) supplemented by six d-type Gaussians for the second-row elements. This basis set therefore attempts to account for d-orbital effects in molecules containing second-row elements. There are no special polarisation functions on first-row elements, which are described by the 3-21G basis set.

A deficiency of the basis sets described so far is their inability to deal with species such as anions and molecules containing lone pairs which have a significant amount of electron density away from the nuclear centres. This failure arises because the amplitudes of the Gaussian basis functions are rather low far from the nuclei. To remedy this deficiency highly diffuse functions can be added to the basis set. These basis sets are denoted using a '+'; thus the 3-21+G basis set contains an additional single set of diffuse s- and p-type Gaussian functions. '++' indicates that the diffuse functions are included for hydrogen as well as for heavy atoms. At these levels the terminology starts to become a little unwieldy. For example, the 6-311++G(3df, 3pd) basis set uses a single zeta core and triple zeta valence representation with additional diffuse functions on all atoms. The '(3df, 3pd)' indicates three sets of d functions and one set of f functions for first-row atoms and three sets of p functions and one set of d functions for hydrogen. This latter convention is probably the most generic; one commonly encountered example is the 6-31G(d) basis set, which is synonymous with 6-31G*.

The basis sets that we have considered thus far are sufficient for most calculations. However, for some high-level calculations a basis set that effectively enables the basis set limit to be achieved is required. The *even-tempered* basis set is designed to achieve this; each function in this basis set is the product of a spherical harmonic and a Gaussian function multiplied

by a power of the distance from the origin:

$$\chi_{klm}(\rho,\theta,\phi) = \exp(-\zeta_k^2) r^l Y_{lm}(\theta,\phi) \qquad (2\ 203)$$

The orbital exponent ζ_k is expressed as a function of two parameters α and β as follows.

$$\zeta_k = \alpha\beta^k \qquad k = 1,2,3,\ldots,N \qquad (2.204)$$

The even-tempered basis set consists of the following sequence of functions: 1s, 2p, 3d, 4f, ..., which correspond to increasing values of k. The advantage of this basis set is that it is relatively easy to optimise the exponents for a large sequence of basis functions.

2.6.1 Creating a Basis Set

There is no definitive method for generating basis sets, and the construction of a new basis set is very much an art. Nevertheless, there are a number of well-established approaches that have resulted in widely used basis sets. We have already seen how linear combinations of Gaussian functions can be fitted to Slater type orbitals by minimising the overlap (see Figure 2.6 and Table 2.3). The Gaussian exponents and coefficients are derived by least-squares fitting to the desired functions, such as Slater type orbitals. When using basis sets that have been fitted to Slater orbitals it is often advantageous to use Slater exponents that are different to those obtained from Slater's rules. In general, better results for molecular calculations are obtained if larger Slater exponents are used for the valence electrons; this has the effect of giving a 'smaller', less diffuse orbital. For example, a value of 1.24 is widely used for the Slater exponent of hydrogen rather than the 1.0 that would be suggested by Slater's rules. It is straightforward to derive a basis set for a different Slater exponent if the Gaussian expansion has been fitted to a Slater type orbital with $\zeta = 1.0$. If the Slater exponent ζ is replaced by a new value, ζ', then the respective Gaussian exponents α and α' are related by:

$$\frac{\alpha'}{\alpha} = \frac{\zeta'^2}{\zeta^2} \qquad (2.205)$$

A doubling of the Slater exponent thus corresponds to a quadrupling of the Gaussian exponent The expansion coefficients remain the same. For example, to obtain the exponents of the Gaussian functions for hydrogen in the STO-3G basis set we need to multiply the appropriate values in Table 2.3 by 1.24^2, giving exponents of 0.168 856, 0.623 913 and 3.425 25. This strategy can be quite powerful; the STO-nG basis sets were originally defined with exponents that reproduce 'best atom' values for the core orbitals, but the exponents for the valence electrons were values that give optimal performance for a selected set of small molecules. For example, the suggested exponent for the valence orbitals in carbon was 1.72 rather than the 1.625 predicted by Slater's rules. The core orbitals have a Slater exponent of 5.67.

Basis sets can be constructed using an optimisation procedure in which the coefficients and the exponents are varied to give the lowest atomic energies. Some complications can arise when this approach is applied to larger basis sets. For example, in an atomic calculation the diffuse functions can move towards the nucleus, especially if the core region is described

by only a few basis functions. This is contrary to the role of diffuse functions, which is to enhance the description in the internuclear region. It may therefore be necessary to construct the basis set in stages, first determining the diffuse functions, using many basis functions for the core, and then optimising the basis functions for the core region, keeping the diffuse functions fixed. In many of the popular Gaussian basis sets the coefficients and exponents of the core orbitals are designed to reproduce calculations on atoms, whereas the valence basis functions are parametrised to reproduce the properties of a carefully selected set of molecular data.

The basis sets of Dunning [Dunning 1970] are obtained in a rather different way to those of Pople and co-workers. The first step is to perform an atomic SCF calculation using a set of primitive Gaussian functions in which the exponents are optimised to give the lowest energy for the atom. This set of primitive Gaussian functions (usually far too many for general use in molecular calculations) is then contracted to a smaller number of Gaussian functions, so drastically reducing the number of integrals that need be calculated. For example, Huzinga optimised the exponents of an uncontracted basis set that contained nine functions of s symmetry and five functions of p symmetry for the first-row elements [Huzinga 1965] This (9s5p) basis set represents the 1s, 2s and three 2p orbitals and in fact corresponds to 24 basis functions per atom ($9 + 3 \times 5$). The primitive Gaussians in this uncontracted basis set are then apportioned to the basis functions in the new, contracted basis set, which contains three s functions and two p functions and is written [3s2p]. No primitive is assigned to more than one of the contracted basis functions. The 1s orbital is constructed from six primitives, the 2s orbital from one set of two primitives and one set containing just one primitive, and the 2p orbitals are represented by one contracted function containing five primitives and one contracted function that contains the remaining primitive. The final basis set, which is illustrated in Table 2.4 for nitrogen, contains a total of nine basis functions rather than the original 24. Each of the primitive functions appears

Exponent	Coefficient	Exponent	Coefficient	Exponent	Coefficient
1s		**2s**		**2s**	
5900	0.001 190	7.193	−0.160 405	0.2133	1 000 000
887.5	0.009 099	1.707	1 058 215		
204.7	0 044 145				
59.84	0 150 464				
20.00	0.356 741				
7.193	0.446 533				
2.686	0 145 603				
2p		**2p**			
26.79	0 018 254	0.1654	1.000 000		
5.956	0.116 461				
1.707	0 390 178				
0.5314	0 637 102				

Table 2.4 Exponents and contraction coefficients for the three s-type and the two p-type Gaussian functions in the basis set of Dunning for nitrogen [Dunning 1970].

in just one basis function with its original exponent. The ratios of the coefficients of the primitives in the contracted basis set are equal to the ratios of the coefficients determined in the atomic SCF calculation. The major advantage of this approach is that calculations with the smaller basis set give results that are almost as good as calculations using the full basis set but with much less computational effort.

2.7 Calculating Molecular Properties Using *ab initio* Quantum Mechanics

We have now considered the key features of the *ab initio* approach to quantum mechanical calculations and so, as an antidote to the rather theoretical nature of the chapter so far, it is appropriate to consider how the method might be used in practice. Quantum mechanics can be used to calculate a wide range of properties. In addition to thermodynamic and structural values, quantum mechanics can be used to derive properties dependent upon the electronic distribution. Such properties often cannot be determined by any other method. In this section we shall provide a flavour of the ways in which quantum mechanics is used in molecular modelling. Other applications, such as the location of transition structures and the use of quantum mechanics in deriving force field parameters, will be discussed in later chapters. Many different computer programs are now available for performing *ab initio* calculations; probably the best known of these is the Gaussian series of programs which originated in the laboratory of John Pople, who has made numerous contributions to the field, recognised by the award of the Nobel Prize in 1998.

2.7.1 Setting Up the Calculation and the Choice of Coordinates

The traditional way to provide the nuclear coordinates to a quantum mechanical program is via a Z-matrix, in which the positions of the nuclei are defined in terms of a set of internal coordinates (see Section 1.2). Some programs also accept coordinates in Cartesian format, which can be more convenient for large systems. It can sometimes be important to choose an appropriate set of internal coordinates, especially when locating minima or transition points or when following reaction pathways. This is discussed in more detail in Section 5.7.

2.7.2 Energies, Koopman's Theorem and Ionisation Potentials

The energy of an electron in an orbital (Equation (2.169)) is often equated with the energy required to remove the electron to give the corresponding ion. This is *Koopman's theorem*. Two important caveats must be remembered when applying Koopman's theorem and comparing the results with experimentally determined ionisation potentials. The first of these is that the orbitals in the ionised state are assumed to be the same as in the unionised state; they are 'frozen'. This neglects the fact that the orbitals in the ionised state will be different from those in the unionised state. The energy of the ionised state will thus tend to be higher than it 'should' be, giving too large an ionisation potential. The second caveat is that the Hartree–Fock method does not include the effects of electron correlation.

The correction due to electron correlation would be expected to be greater for the unionised state than for the ionised state, as the former has more electrons. Fortunately, therefore, the effect of electron correlation often opposes the effect of the frozen orbitals, resulting in many cases in good agreement between experimentally determined ionisation potentials and calculated values.

A Hartree–Fock SCF calculation with K basis functions provides K molecular orbitals, but many of these will not be occupied by any electrons; they are the 'virtual' spin orbitals. If we were to add an electron to one of these virtual orbitals then this should provide a means of calculating the electron affinity of the system. Electron affinities predicted by Koopman's theorem are always positive when Hartree–Fock calculations are used, because the virtual orbitals always have a positive energy. However, it is observed experimentally that many neutral molecules will accept an electron to form a stable anion and so have negative electron affinities. This can be understood if one realises that electron correlation would be expected to add to the error due to the 'frozen' orbital approximation, rather than to counteract it as for ionisation potentials.

2.7.3 Calculation of Electric Multipoles

Some of the most important properties that a quantum mechanical calculation provides are the electric multipole moments of the molecule. The electric multipoles reflect the distribution of charge in a molecule. The simplest electric moment (apart from the total net charge on the molecule) is the dipole. The dipole moment of a distribution of charges q_i located at positions $\mathbf{r}_i$ is given by $\sum q_i \mathbf{r}_i$. If there are just two charges $+q$ and $-q$ separated by a distance r then the dipole moment is qr. A dipole moment of 4.8 Debye corresponds to two charges equal in magnitude to the electronic charge e separated by 1 Å. The dipole moment is a vector quantity, with components along the three Cartesian axes. The dipole moment of a molecule has contributions from both the nuclei and the electrons. The nuclear contributions can be calculated using the formula for a system of discrete charges:

$$\boldsymbol{\mu}_{\text{nuclear}} = \sum_{A=1}^{M} Z_A \mathbf{R}_A \tag{2.206}$$

The electronic contribution arises from a continuous function of electron density and must be calculated using the appropriate operator:

$$\boldsymbol{\mu}_{\text{electronic}} = \int d\tau \Psi_0 \left(\sum_{i=1}^{N} -\mathbf{r}_i \right) \Psi_0 \tag{2.207}$$

The dipole moment operator is a sum of one-electron operators $\mathbf{r}_i$, and as such the electronic contribution to the dipole moment can be written as a sum of one-electron contributions. The electronic contribution can also be written in terms of the density matrix, $\mathbf{P}$, as follows:

$$\boldsymbol{\mu}_{\text{electronic}} = \sum_{\mu=1}^{K} \sum_{\nu=1}^{K} P_{\mu\nu} \int d\tau \phi_\mu(-\mathbf{r}) \phi_\nu \tag{2.208}$$

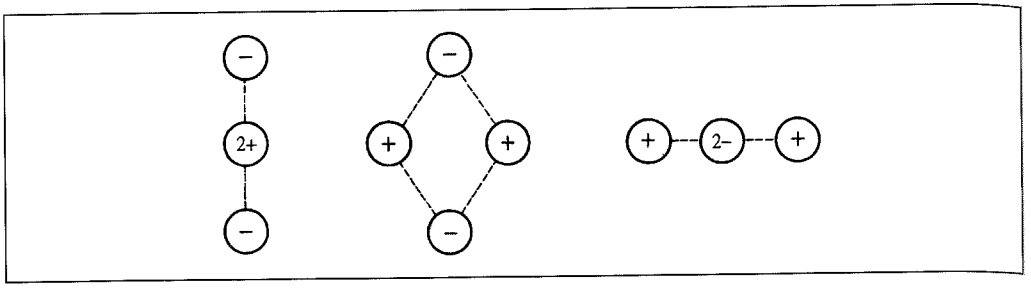

Fig. 2 8. A quadrupole moment can be obtained from various arrangements of two positive and two negative charges

The electronic contribution to the dipole moment is thus determined from the density matrix and a series of one-electron integrals $\int d\tau \phi_\mu(-\mathbf{r})\phi_\nu$. The dipole moment operator, r, has components in the x, y and z directions, and so these one-electron integrals are divided into their appropriate components; for example, the x component of the electronic contribution to the dipole moment would be determined using:

$$\mu_x = \sum_{\mu=1}^{K} \sum_{\nu=1}^{K} P_{\mu\nu} \int d\tau \phi_\mu(-x)\phi_\nu \tag{2.209}$$

The quadrupole is the next electric moment. A molecule has a non-zero electric quadrupole moment when there is a non-spherically symmetrical distribution of charge. A quadrupole can be considered to arise from four charges that sum to zero which are arranged so that they do not lead to a net dipole. Three such arrangements are shown in Figure 2.8. Whereas the dipole moment has components in the x, y and z directions, the quadrupole has nine components from all pairwise combinations of x and y and is represented by a 3×3 matrix as follows:

$$\Theta = \begin{pmatrix} \sum q_i x_i^2 & \sum q_i x_i y_i & \sum q_i x_i z_i \\ \sum q_i y_i x_i & \sum q_i y_i^2 & \sum q_i y_i z_i \\ \sum q_i z_i x_i & \sum q_i z_i y_i & \sum q_i z_i^2 \end{pmatrix} \tag{2 210}$$

The three moments higher than the quadrupole are the octopole, hexapole and decapole. Methane is an example of a molecule whose lowest non-zero multipole moment is the octopole. The entire set of electric moments is required to completely and exactly describe the distribution of charge in a molecule. However, the series expansion is often truncated after the dipole or quadrupole as these are often the most significant.

Extensive comparisons have been made of experimental and calculated dipole moments (and in some cases the higher moments, though these are difficult to determine accurately by experiment). Factors such as the basis set and electron correlation can have a significant impact on the accuracy of the results, but it is found in many cases that the errors are systematic and that a simple scaling factor can be used to convert the results of a calculation with a small basis set to those obtained from experiment or with a much larger basis set. To illustrate how calculated dipole moments can vary, Table 2.5 provides the dipole moments for formaldehyde calculated at the experimental geometry using a variety of basis sets. It is

STO-3G	1.5258	3-21G	2.2903	4-31G	3.0041
6-31G*	2.7600	6-31G**	2.7576	6-311G**	2.7807
Expt.	2 34				

Table 2 5 Dipole moments calculated for formaldehyde using various basis sets at the experimental geometry.

also important to note that the dipole moment can be very sensitive to the geometry from which it is calculated.

2.7.4 The Total Electron Density Distribution and Molecular Orbitals

The electron density $\rho(\mathbf{r})$ at a point $\mathbf{r}$ can be calculated from the Born interpretation of the wavefunction as a sum of squares of the spin orbitals at the point $\mathbf{r}$ for all occupied molecular orbitals. For a system of N electrons occupying $N/2$ real orbitals, we can write:

$$\rho(\mathbf{r}) = 2\sum_{i=1}^{N/2} |\psi_i(\mathbf{r})|^2 \tag{2.211}$$

If we express the molecular orbital ψ_i as a linear combination of basis functions, then the electron density at a point $\mathbf{r}$ is given as:

$$\begin{aligned}\rho(\mathbf{r}) &= 2\sum_{i=1}^{N/2}\left(\sum_{\mu=1}^{K} c_{\mu i}\phi_\mu(\mathbf{r})\right)\left(\sum_{\nu=1}^{K} c_{\nu i}\phi_\nu(\mathbf{r})\right) \\ &= 2\sum_{i=1}^{N/2}\sum_{\mu=1}^{K} c_{\mu i}c_{\mu i}\phi_\mu(\mathbf{r})\phi_\mu(\mathbf{r}) + 2\sum_{i=1}^{N/2}\sum_{\mu=1}^{K}\sum_{\nu=\mu+1}^{K} 2c_{\mu i}c_{\nu i}\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})\end{aligned} \tag{2 212}$$

Equation (2.212) can be tidied up considerably if it is written in terms of the elements of the density matrix:

$$\left(P_{\mu\nu} = 2\sum_{i=1}^{N/2} c_{\mu i}c_{\nu i}\right)$$

$$\begin{aligned}\rho(\mathbf{r}) &= \sum_{\mu=1}^{K}\sum_{\nu=1}^{K} P_{\mu\nu}\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r}) \\ &= \sum_{\mu=1}^{K} P_{\mu\mu}\phi_\mu(\mathbf{r})\phi_\mu(\mathbf{r}) + 2\sum_{\mu=1}^{K}\sum_{\nu=\mu+1}^{K} P_{\mu\nu}\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})\end{aligned} \tag{2.213}$$

The integral of $\rho(\mathbf{r})$ over all space equals the number of electrons in the system, N:

$$N = \int d\mathbf{r}\rho(\mathbf{r}) = 2\sum_{i=1}^{N/2}\int d\mathbf{r}|\psi_i(\mathbf{r})|^2 \tag{2.214}$$

If the overlap between two orbitals ϕ_μ and ϕ_ν is written as $S_{\mu\nu}$, and if the basis functions are

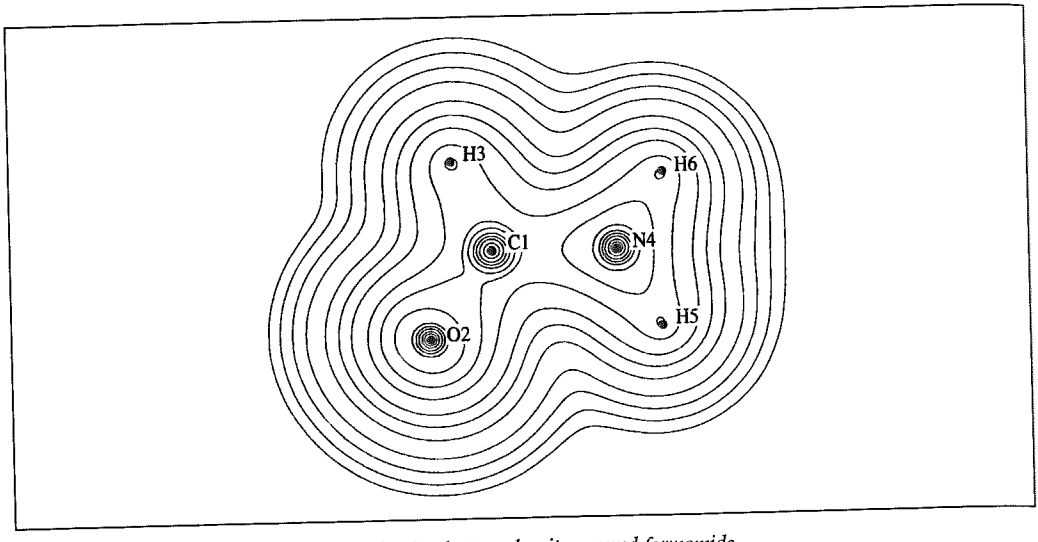

Fig. 2 9: Contour map showing the variation in electron density around formamide

assumed to be normalised ($S_{\mu\mu} = 1$), then:

$$N = \sum_{\mu=1}^{K} P_{\mu\mu} + 2\sum_{\mu=1}^{K}\sum_{\nu=\mu+1}^{K} P_{\mu\nu}S_{\mu\nu} \qquad (2\ 215)$$

The electron density can be visualised in several ways. One approach is to construct contours on slices through the molecule, such that each contour connects points of equal density, as shown in Figure 2.9 for formamide. The electron density can also be represented as an isometric projection (or a 'relief map', Figure 2.10), in which the height above the plane

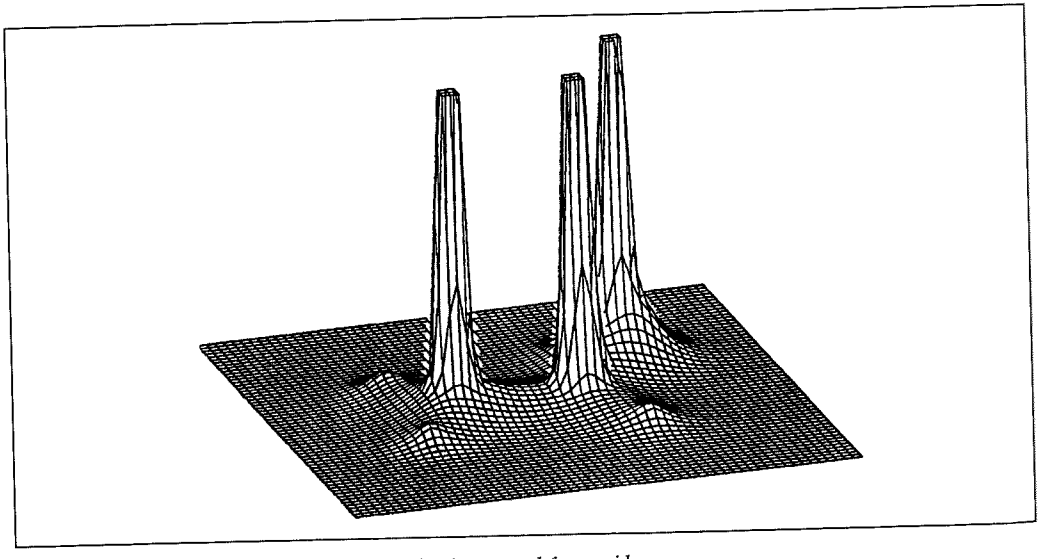

Fig 2 10: Isometric projection of the electron density around formamide

represents the magnitude of the electron density. These diagrams show that the electron density tends to be greatest near the nuclei, as would be expected. The electron density can also be represented as a solid object, whose surface connects points of equal density. The surface shown in Figure 2.11 (colour plate section) corresponds to an electron density of 0.0001 a.u. around formamide. Other properties such as the electrostatic potential can be mapped onto this surface, as we shall see in Section 2.7.9.

The electron density distribution of individual molecular orbitals may also be determined and plotted The highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) are often of particular interest as these are the orbitals most commonly involved in chemical reactions. As an illustration, the HOMO and LUMO for formamide are displayed in Figures 2.12 and 2.13 (colour plate section) as surface pictures.

2.7.5 Population Analysis

Population analysis methods partition the electron density between the nuclei so that each nucleus has a 'number' (not necessarily an integral number) of electrons associated with it. Such a partitioning provides a way to calculate the atomic charge on each nucleus. It should be noted that there is no quantum mechanical operator for the atomic charge and so any partitioning scheme must be arbitrary. Hence many methods have been devised. Here we will consider Mulliken and Löwdin analysis and Bader's theory of atoms in molecules. The alternatives include natural population analysis [Reed *et al.* 1985; Bachrach 1994]. Wiberg and Rablen have compared a number of methods for calculating atomic charges, and we refer to some of their results in the following discussion [Wiberg and Rablen 1993]. To illustrate the variation that can be obtained in the results, for methane they found that the charge on the carbon atom varied from −0.473 to +0.244, depending upon the method chosen! We will also consider the problem of calculating atomic charges in more detail in Chapter 4 on molecular mechanics.

2.7.6 Mulliken and Löwdin Population Analysis

R S Mulliken suggested a widely used method for performing population analysis [Mulliken 1955]. The starting point is Equation (2.215), which relates the total number of electrons to the density matrix and to the overlap integrals. In the Mulliken method, all of the electron density ($P_{\mu\mu}$) in an orbital is allocated to the atom on which ϕ_μ is located. The remaining electron density is associated with the overlap population, $\phi_\mu\phi_\nu$. For each element $\phi_\mu\phi_\nu$ of the density matrix, half of the density is assigned to the atom on which ϕ_μ is located and half to the atom on which ϕ_ν is located. The net charge on an atom A is then calculated by subtracting the number of electrons from the nuclear charge, Z_A:

$$q_A = Z_A - \sum_{\mu=1;\,\mu \text{ on A}}^{K} P_{\mu\mu} - \sum_{\mu=1;\,\mu \text{ on A}}^{K} \sum_{\nu=1;\,\nu\neq\mu}^{K} P_{\mu\nu}S_{\mu\nu} \qquad (2.216)$$

Mulliken population analysis is a trivial calculation to perform once a self-consistent field has been established and the elements of the density matrix have been determined

However, there are some serious shortcomings to the method, as Mulliken himself pointed out.

A Mulliken analysis depends upon the use of a balanced basis set, in which an equivalent number of basis functions is present on each atom in the molecule. For example, it is possible to calculate a wavefunction for a molecule such as water in which all of the basis functions reside on the oxygen atom; if a large enough basis set is used then a quite reasonable wavefunction for the whole molecule can be obtained. However, the Mulliken analysis would put all of the charge on the oxygen. This is an extreme example of a general problem; p, d and f orbitals are spread quite far from the nucleus with which they are associated and so may be very close to other atoms, yet the charge associated with electron occupation of such orbitals is assigned to the atom on which the orbital is centred. The equal apportioning of electrons between pairs of atoms, even if their electronegativities are very different, can lead in some cases to quite unrealistic values for the net atomic charge. *In extremis*, some orbitals may 'contain' a negative number of electrons and others more than two electrons, in clear contradiction of the Pauli principle. A Mulliken analysis assumes that each basis function can be associated with an atomic centre and so is not applicable if basis functions not centred on the nuclei are used. The atomic charges can be very dependent upon the basis set; for example, Wiberg and Rablen found that the charge on the central carbon in isobutene changed from +0.1 with a 6-31G* basis set to +1.0 for a 6-311++G** basis set.

In the Löwdin approach to population analysis [Löwdin 1970; Cusachs and Politzer 1968] the atomic orbitals are transformed to an orthogonal set, along with the molecular orbital coefficients. The transformed orbitals ϕ'_μ in the orthogonal set are given by:

$$\phi'_\mu = \sum_{\nu=1}^{K} (\mathbf{S}^{-1/2})_{\nu\mu}\phi_\nu \tag{2.217}$$

The electron population associated with an atom becomes:

$$q_A = Z_A - \sum_{\mu=1,\,\mu \text{ on A}}^{K} (\mathbf{S}^{1/2}\mathbf{P}\mathbf{S}^{1/2})_{\mu\mu} \tag{2.218}$$

Löwdin population analysis avoids the problem of negative populations or populations greater than 2. Some quantum chemists prefer the Löwdin approach to that of Mulliken as the charges are often closer to chemically intuitive values and are less sensitive to basis set.

2.7.7 Partitioning Electron Density: The Theory of Atoms in Molecules

R F W Bader's theory of 'atoms in molecules' [Bader 1985] provides an alternative way to partition the electrons between the atoms in a molecule. Bader's theory has been applied to many different problems, but for the purposes of our present discussion we will concentrate on its use in partitioning electron density. The Bader approach is based upon the concept of a *gradient vector path*, which is a curve around the molecule such that it is always perpendicular to the electron density contours. A set of gradient paths is drawn in Figure 2 14 for formamide. As can be seen, some of the gradient paths terminate at the atomic nuclei. Other gradient paths are attracted to points (called critical points) that are

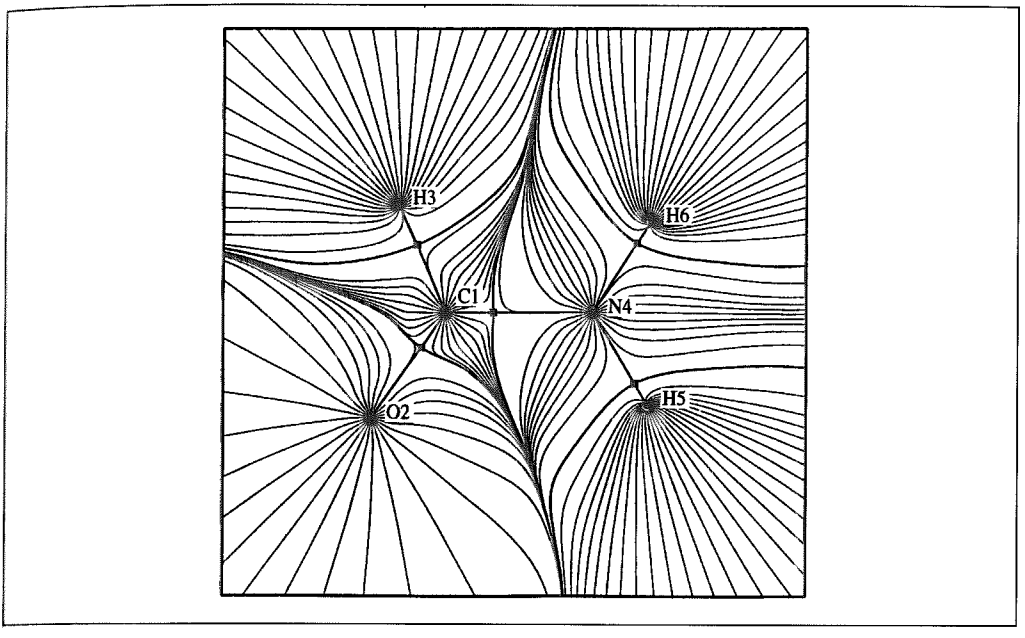

Fig 2 14: Gradient vector paths around formamide The paths terminate at atoms or at bond critical points (indicated by squares)

not located at the nuclei; particularly common are the bond critical points, which are located between bonded atoms. Other types of critical point can occur; for example, a *ring critical point* is found in the centre of a benzene ring.

The bond critical points are points of minimum electron charge density between two bonded atoms. If we follow the contour in three-dimensional space from such a point down the gradient path along which the density decreases most rapidly then this gives a means of partitioning the density. This is shown in Figure 2.15 for hydrogen fluoride and in Figure 2.16 for formamide. This procedure can be performed for each bond, resulting in a three-dimensional partitioning of the electron density. The electron population that is assigned to each atom is then calculated by numerically integrating the charge density within the region surrounding that atom.

Wiberg and Rablen found that the charges obtained with the atoms in molecules method were relatively invariant to the basis set. The charges from this method were also consistent with the experimentally determined C–H bond dipoles in methane (in which the carbon is positive) and ethyne (in which the carbon is negative), unlike most of the other methods they examined.

2.7.8 Bond Orders

As with atomic charges, the bond order is not a quantum mechanical observable and so various methods have been proposed for calculating the bond orders in a molecule.

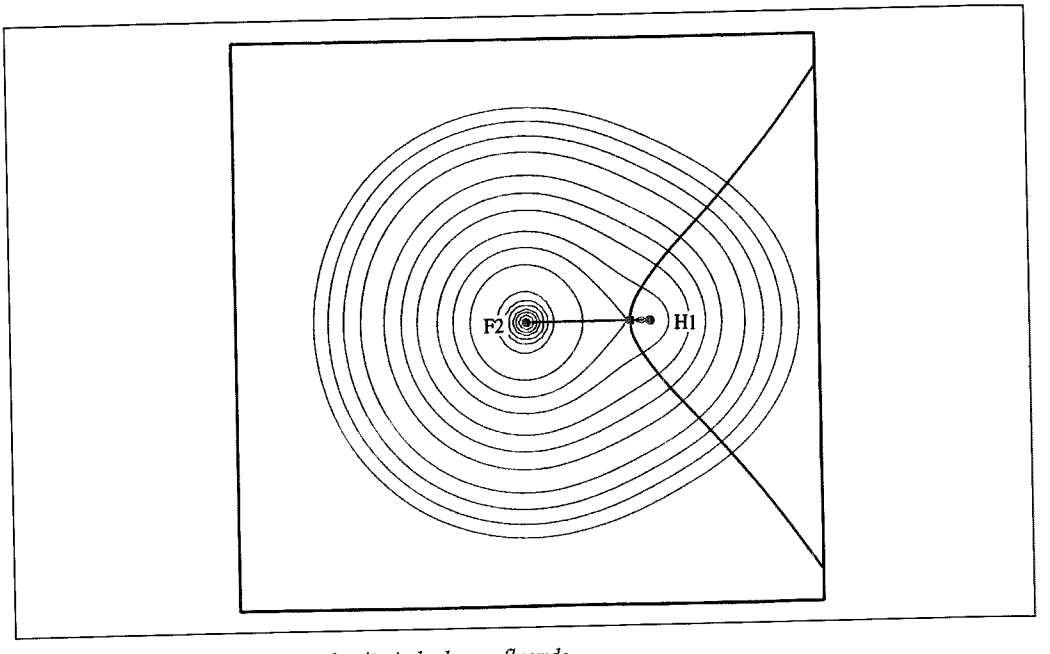

Fig 2 15· Partitioning the electron density in hydrogen fluoride

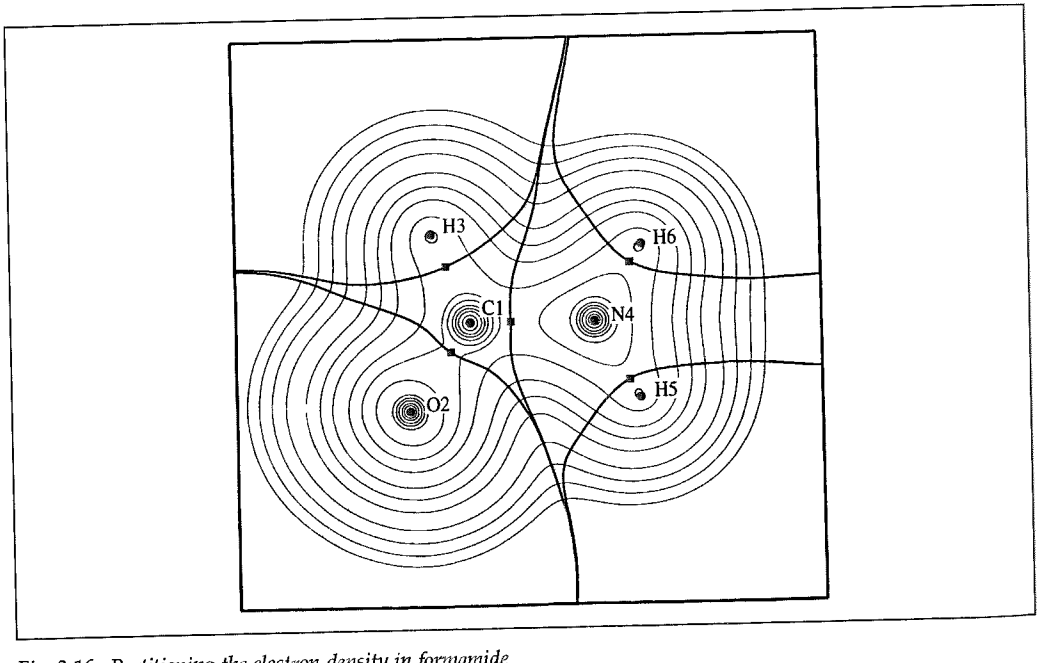

Fig 2 16· Partitioning the electron density in formamide

Molecule	Bond	STO-3G	4-31G
H_2	H–H	1.0	1.0
Methane	C–H	0.99	0 96
Ethene	C=C	2.01	1.96
	C–H	0 98	0.96
Ethyne	C≡C	3 00	3.27
	C–H	0.98	0.86
Water	O–H	0.95	0.80
N_2	N≡N	3.0	2.67

Table 2.6 Bond order obtained from the Mayer bond order scheme [Mayer 1983]

Mayer defined the bond order between two atoms as follows [Mayer 1983]:

$$B_{AB} = \sum_{\mu \text{ on A}} \sum_{\nu \text{ on B}} [(\mathbf{PS})_{\mu\nu}(\mathbf{PS})_{\nu\mu} + (\mathbf{P}^s\mathbf{S})_{\mu\nu}(\mathbf{P}^s\mathbf{S})_{\nu\mu}] \tag{2.219}$$

$\mathbf{P}$ is the total spinless density matrix ($\mathbf{P} = \mathbf{P}^\alpha + \mathbf{P}^\beta$) and $\mathbf{P}^s$ is the spin density matrix ($\mathbf{P}^s = \mathbf{P}^\alpha + \mathbf{P}^\beta$). For a closed-shell system Mayer's definition of the bond order reduces to:

$$B_{AB} = \sum_{\mu \text{ on A}} \sum_{\nu \text{ on B}} (\mathbf{PS})_{\mu\nu}(\mathbf{PS})_{\nu\mu} \tag{2.220}$$

The bond orders obtained from Mayer's formula often seem intuitively reasonable, as illustrated in Table 2.6 for some simple molecules. The method has also been used to compute the bond orders for intermediate structures in reactions of the form $H + XH \rightarrow HX + H$ and $X + H_2 \rightarrow XH + H$ (X = F, Cl, Br). The results suggested that bond orders were a useful way to describe the similarity of the transition structure to the reactants or to the products. Moreover, the bond orders were approximately conserved along the reaction pathway.

As with methods for allocating electron density to atoms, the Mayer method is not necessarily 'correct', though it appears to be a useful measure of the bond order that conforms to accepted pictures of bonding in molecules.

2.7.9 Electrostatic Potentials

The electrostatic potential at a point $\mathbf{r}$, $\phi(\mathbf{r})$, is defined as the work done to bring unit positive charge from infinity to the point. The electrostatic interaction energy between a point charge q located at $\mathbf{r}$ and the molecule equals $q\phi(\mathbf{r})$. The electrostatic potential has contributions from both the nuclei and from the electrons, unlike the electron density, which only reflects the electronic distribution. The electrostatic potential due to the M nuclei is:

$$\phi_{\text{nucl}}(\mathbf{r}) = \sum_{A=1}^{M} \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \tag{2.221}$$

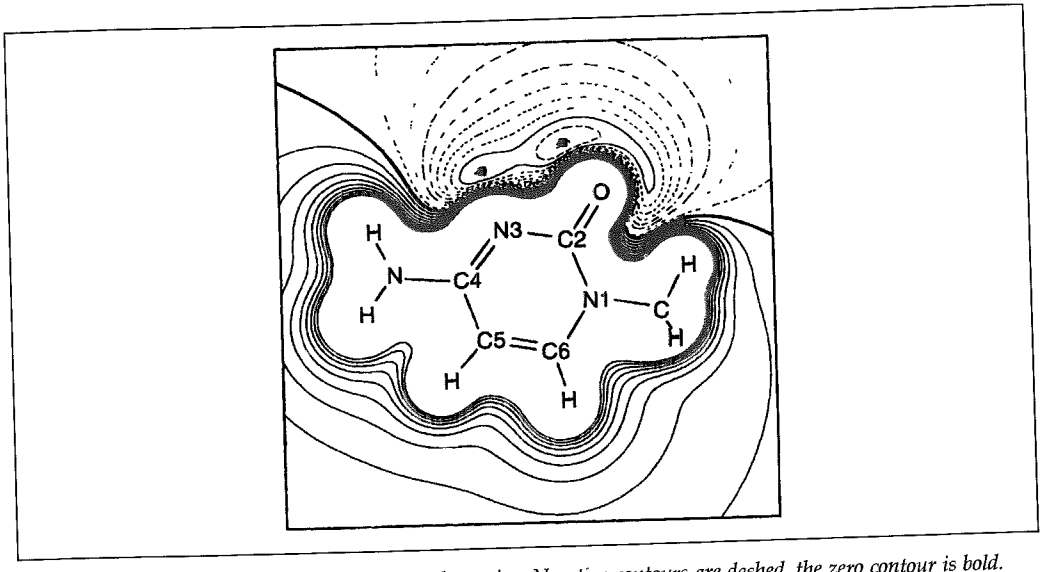

Fig 2 17: Electrostatic potential contours around cytosine Negative contours are dashed, the zero contour is bold. The minima near N3 and O are marked

The potential due to the electrons is obtained from the appropriate integral of the electron density:

$$\phi_{\text{elec}}(\mathbf{r}) = -\int \frac{d\mathbf{r}'\rho(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} \tag{2.222}$$

The total electrostatic potential equals the sum of the nuclear and the electronic contributions:

$$\phi(\mathbf{r}) = \phi_{\text{nucl}}(\mathbf{r}) + \phi_{\text{elec}}(\mathbf{r}) \tag{2.223}$$

The electrostatic potential has proved to be particularly useful for rationalising the interactions between molecules and molecular recognition processes. This is because electrostatic forces are primarily responsible for long-range interactions between molecules. The electrostatic potential varies through space, and so it can be calculated and visualised in the same way as the electron density. Electrostatic potential contours can be used to propose where electrophilic attack might occur; electrophiles are often attracted to regions where the electrostatic potential is most negative. For example, the experimentally determined position of electrophilic attack at the nucleic acid cytosine is at N3 (Figure 2.17). This atom is next to a minimum in the electrostatic potential (also shown in Figure 2.17), as pointed out by Politzer and Murray [Politzer and Murray 1991].

Non-covalent interactions between molecules often occur at separations where the van der Waals radii of the atoms are just touching and so it is often most useful to examine the electrostatic potential in this region. For this reason, the electrostatic potential is often calculated at the molecular surface (defined in Section 1.5) or the equivalent isodensity surface as shown in Figure 2.18 (colour plate section). Such pictorial representations

can be used to qualitatively assess the degree of electrostatic similarity between two molecules.

2.7.10 Thermodynamic and Structural Properties

The total energy of a system is equal to the sum of the electronic energy and the Coulombic nuclear repulsion energy:

$$E_{\text{tot}} = E_{\text{elec}} + \sum_{A=1}^{M} \sum_{B=A+1}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{2.224}$$

A more useful quantity for comparison with experiment is the heat of formation, which is defined as the enthalpy change when one mole of a compound is formed from its constituent elements in their standard states. The heat of formation can thus be calculated by subtracting the heats of atomisation of the elements and the atomic ionisation energies from the total energy. Unfortunately, *ab initio* calculations that do not include electron correlation (which we will discuss in Chapter 3) provide uniformly poor estimates of heats of formation with errors in bond dissociation energies of 25–40 kcal/mol, even at the Hartree–Fock limit for diatomic molecules.

When combined with an energy minimisation algorithm, quantum mechanics can be used to calculate equilibrium geometries of molecules. The results of such calculations can be compared with the structures obtained from gas-phase experiments using microwave spectroscopy, electronic spectroscopy and electron diffraction. Extensive tables listing comparisons between calculations and experiment for many molecules have been published in several reviews. Not surprisingly, the agreement between theory and experiment for *ab initio* calculations generally improves as one increases the size of the basis set. Hehre *et al.* suggest that the 3-21G basis set offers a good compromise between performance and applicability [Hehre *et al.* 1986]. It is often found that errors in structural predictions are systematic rather than random. For example, STO-3G bond lengths are generally too long, whilst 6-31G* bond lengths tend to be too short. By analysing the trends in such calculations it can be possible to derive scaling factors which enable more accurate predictions to be made for each level of theory.

Quantum mechanics can be used to calculate the relative energies of conformations and the energy barriers between them. Experimental data is available for both relative stabilities and barrier heights in some cases, though this tends to be limited to relatively simple molecules. Butane is one molecule that has been investigated in great detail, with its *gauche* and *anti* conformations and the barriers that separate them. The energy difference between the *syn* and *anti* conformations of butane (Figure 2.19) was found to fall significantly with increasing basis set size, particularly when correlated levels of theory were employed [Wiberg and Murcko 1988; Allinger *et al.* 1990; Smith and Jaffe 1996]. However, the smaller energy difference between the minimum energy *anti* and *gauche* conformations can be calculated quite accurately even with a relatively small basis set. Quantum mechanics calculations of the change in energy as a bond is rotated are often used to parametrise the torsional terms in molecular mechanics force fields, as will be discussed in Section 4.18.

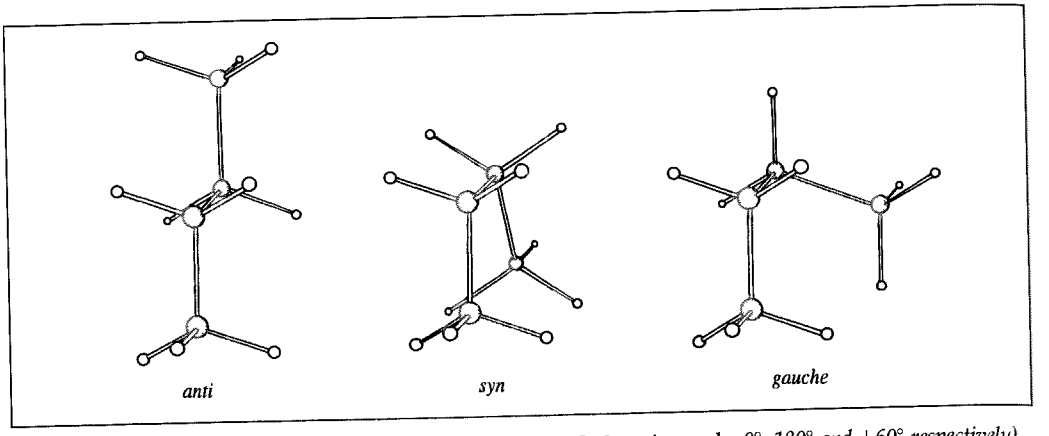

Fig 2.19. syn, anti *and* gauche *conformations of butane (C–C–C–C torsion angles 0°, 180° and ±60° respectively).*

2.8 Approximate Molecular Orbital Theories

Ab initio calculations can be extremely expensive in terms of the computer resources required. Nevertheless, improvements in computer hardware and the availability of easy-to-use programs have helped to make *ab initio* methods a widely used computational tool. The approximate quantum mechanical methods require significantly less computational resources. Indeed, the earliest approximate methods such as Hückel theory predate computers by many years. Moreover, by their incorporation of parameters derived from experimental data some approximate methods can calculate certain properties more accurately then even the highest level of *ab initio* methods.

Many approximate molecular orbital theories have been devised. Most of these methods are not in widespread use today in their original form. Nevertheless, the more widely used methods of today are derived from earlier formalisms, which we will therefore consider where appropriate. We will concentrate on the semi-empirical methods developed in the research groups of Pople and Dewar. The former pioneered the CNDO, INDO and NDDO methods, which are now relatively little used in their original form but provided the basis for subsequent work by the Dewar group, whose research resulted in the popular MINDO/3, MNDO and AM1 methods. Our aim will be to show how the theory can be applied in a practical way, not only to highlight their successes but also to show where problems were encountered and how these problems were overcome. We will also consider the Hückel molecular orbital approach and the extended Hückel method Our discussion of the underlying theoretical background of the approximate molecular orbital methods will be based on the Roothaan–Hall framework we have already developed. This will help us to establish the similarities and the differences with the *ab initio* approach.

2.9 Semi-empirical Methods

A discussion of semi-empirical methods starts most appropriately with the key components

of the Roothaan–Hall equations, which for a closed-shell system are:

$$\mathbf{FC} = \mathbf{SCE} \tag{2.225}$$

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + \sum_{\lambda=1}^{K}\sum_{\sigma=1}^{K} P_{\lambda\sigma}[(\mu\nu|\lambda\sigma) - \tfrac{1}{2}(\mu\lambda|\nu\sigma)] \tag{2.226}$$

$$P_{\lambda\sigma} = 2\sum_{i=1}^{N/2} c_{\lambda i}c_{\sigma i} \tag{2.227}$$

$$H_{\mu\nu}^{\text{core}} = \int d\nu_1 \phi_\mu(1)\left[-\frac{1}{2}\nabla^2 - \sum_{A=1}^{M}\frac{Z_A}{|r_1 - R_A|}\right]\phi_\nu(1) \tag{2.228}$$

In *ab initio* calculations all elements of the Fock matrix are calculated using Equation (2.226), irrespective of whether the basis functions ϕ_μ, ϕ_ν, ϕ_λ and ϕ_σ are on the same atom, on atoms that are bonded or on atoms that are not formally bonded. To discuss the semi-empirical methods it is useful to consider the Fock matrix elements in three groups: $F_{\mu\mu}$ (the diagonal elements), $F_{\mu\nu}$ (where ϕ_μ and ϕ_ν are on the same atom) and $F_{\mu\nu}$ (where ϕ_μ and ϕ_ν are on different atoms).

We have mentioned several times that the greatest proportion of the time required to perform an *ab initio* Hartree–Fock SCF calculation is invariably spent calculating and manipulating integrals. The most obvious way to reduce the computational effort is therefore to neglect or approximate some of these integrals. Semi-empirical methods achieve this in part by explicitly considering only the valence electrons of the system; the core electrons are subsumed into the nuclear core. The rationale behind this approximation is that the electrons involved in chemical bonding and other phenomena that we might wish to investigate are those in the valence shell. By considering all the valence electrons the semi-empirical methods differ from those theories (e.g. Hückel theory) that explicitly consider only the π electrons of a conjugated system and which are therefore limited to specific classes of molecule. The semi-empirical calculations invariably use basis sets comprising Slater type s, p and sometimes d orbitals. The orthogonality of such orbitals enables further simplifications to be made to the equations.

A feature common to the semi-empirical methods is that the overlap matrix, **S** (in Equation (2.225)), is set equal to the identity matrix I. Thus all diagonal elements of the overlap matrix are equal to 1 and all off-diagonal elements are zero. Some of the off-diagonal elements would naturally be zero due to the use of orthogonal basis sets on each atom, but in addition the elements that correspond to the overlap between two atomic orbitals on different atoms are also set to zero. The main implication of this is that the Roothaan–Hall equations are simplified: $\mathbf{FC} = \mathbf{SCE}$ becomes $\mathbf{FC} = \mathbf{CE}$ and so is immediately in standard matrix form. It is important to note that setting **S** equal to the identity matrix does not mean that all overlap integrals are set to zero in the calculation of Fock matrix elements. Indeed, it is important specifically to include some of the overlaps in even the simplest of the semi-empirical models.

2.9.1 Zero-differential Overlap

Many semi-empirical theories are based upon the zero-differential overlap approximation (ZDO). In this approximation, the overlap between pairs of different orbitals is set to zero for all volume elements dv:

$$\phi_\mu \phi_\nu \, dv = 0 \tag{2.229}$$

This directly leads to the following result for the overlap integrals:

$$S_{\mu\nu} = \delta_{\mu\nu} \tag{2.230}$$

If the two atomic orbitals ϕ_μ and ϕ_ν are located on different atoms then the differential overlap is referred to as diatomic differential overlap; if ϕ_μ and ϕ_ν are on the same atom then we have monatomic differential overlap. If the ZDO approximation is applied to the two-electron repulsion integral $(\mu\nu|\lambda\sigma)$ then the integral will equal zero if $\mu \neq \nu$ and/or if $\lambda \neq \sigma$. This can be written concisely using the Kronecker delta:

$$(\mu\nu|\lambda\sigma) = (\mu\mu|\lambda\lambda)\delta_{\mu\nu}\delta_{\lambda\sigma} \tag{2.231}$$

It can immediately be seen that all three- and four-centre integrals are set to zero under the ZDO approximation. If the ZDO approximation is applied to all orbital pairs then the Roothaan–Hall equations for a closed-shell molecule (Equation (2.226)) simplify considerably to give the following for $\mu \equiv \nu$:

$$F_{\mu\mu} = H_{\mu\mu}^{\text{core}} + \sum_{\lambda=1}^{K} P_{\lambda\lambda}(\mu\mu|\lambda\lambda) - \tfrac{1}{2} P_{\mu\mu}(\mu\mu|\mu\mu) \tag{2.232}$$

The summation over λ includes $\lambda = \mu$, and the terms in $(\mu\mu|\mu\mu)$ can be separated to give:

$$F_{\mu\mu} = H_{\mu\mu}^{\text{core}} + \tfrac{1}{2} P_{\mu\mu}(\mu\mu|\mu\mu) + \sum_{\lambda=1;\, \lambda\neq\mu}^{K} P_{\lambda\lambda}(\mu\mu|\lambda\lambda) \tag{2.233}$$

For $\nu \neq \mu$ we have.

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} - \tfrac{1}{2} P_{\mu\nu}(\mu\mu|\nu\nu) \tag{2.234}$$

Sensible results cannot be obtained by simply applying the ZDO approximation to all pairs of orbitals *carte blanche*. There are two major reasons for this.

The first consideration is that the total wavefunction and the molecular properties calculated from it should be the same when a transformed basis set is used. We have already encountered this requirement in our discussion of the transformation of the Roothaan–Hall equations to an orthogonal set. To reiterate: suppose a molecular orbital is written as a linear combination of atomic orbitals:

$$\psi_i = \sum_\mu c_{\mu i} \phi_\mu \tag{2 235}$$

If an alternative basis set is used in which the basis functions are just linear combinations of the original basis functions, then the same wavefunction can be written as a linear

combination of these new transformed functions:

$$\psi_i = \sum_{\alpha} c_{\alpha i} \phi'_{\alpha} \tag{2.236}$$

$$\phi'_{\alpha} = \sum_{\mu_{\alpha}} t_{\mu_{\alpha}} \phi_{\mu} \tag{2.237}$$

t_{μ_α} are the coefficients of the original basis functions in the linear expansion of the transformed basis set. Different types of transformation are possible; for example, some transformations mix orbitals with the same principal and azimuthal quantum numbers (e.g. mixing $2p_x$, $2p_y$ and $2p_z$); others mix orbitals with the same principal quantum number but different azimuthal quantum numbers (e.g. mixing 2s, $2p_x$, $2p_y$ and $2p_z$ orbitals to give sp^3 hybrid orbitals); yet other transformations mix orbitals located on different atoms. Suppose we mix $2p_x$ and $2p_y$ atomic orbitals on the same atom. The differential overlap between these two orbitals is $2p_x 2p_y$. We now introduce the following two new coordinates, which correspond to a rotation in the xy plane:

$$x' = \frac{1}{\sqrt{2}}(x + y) \tag{2 238}$$

$$y' = \frac{1}{\sqrt{2}}(-x + y) \tag{2.239}$$

The overlap between the $2p'_x$ and $2p'_y$ orbitals in this new coordinate system is $\frac{1}{2}(2p_y^2 - 2p_x^2)$. If the zero differential overlap approximation were applied, then different results would be obtained for the two coordinate systems unless the overlap in the new, transformed system was also ignored.

The second reason why the ZDO approximation is not applied to all pairs of orbitals is that the major contributors to bond formation are the electron–core interactions between pairs of orbitals and the nuclear cores (i.e. $H_{\mu\nu}^{\text{core}}$). These interactions are therefore not subjected to the ZDO approximation (and so do not suffer from any transformation problems).

2.9.2 CNDO

The complete neglect of differential overlap (CNDO) approach of Pople, Santry and Segal was the first method to implement the zero-differential overlap approximation in a practical fashion [Pople *et al.* 1965]. To overcome the problems of rotational invariance, the two-electron integrals $(\mu\mu|\lambda\lambda)$, where μ and λ are on different atoms A and B, were set equal to a parameter γ_{AB} which depends only on the nature of the atoms A and B and the internuclear distance, and not on the type of orbital. The parameter γ_{AB} can be considered to be the average electrostatic repulsion between an electron on atom A and an electron on atom B. When both atomic orbitals are on the same atom the parameter is written γ_{AA} and represents the average electron–electron repulsion between two electrons on an atom A.

With this approximation we can divide the elements of the Fock matrix into three groups: $F_{\mu\mu}$ (the diagonal elements), $F_{\mu\nu}$ (where μ and ν are on different atoms) and $F_{\mu\nu}$ (where μ

and ν are on the same atom). To obtain $F_{\mu\mu}$ we substitute γ_{AB} for the two-electron integrals $(\mu\mu|\lambda\lambda)$ where μ and λ are on different atoms and γ_{AA} where μ and λ are on the same atom into the Fock matrix equations, Equations (2.240)–(2.242):

$$F_{\mu\mu} = H_{\mu\mu}^{\text{core}} + \sum_{\lambda=1;\ \lambda \text{ on A}}^{K} P_{\lambda\lambda}\gamma_{AA} - \tfrac{1}{2}P_{\mu\mu}\gamma_{AA} + \sum_{\lambda=1;\ \lambda \text{ not on A}}^{K} P_{\lambda\lambda}\gamma_{AB} \tag{2.240}$$

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} - \tfrac{1}{2}P_{\mu\nu}\gamma_{AA}; \quad \mu \text{ and } \nu \text{ both on atom A} \tag{2.241}$$

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} - \tfrac{1}{2}P_{\mu\nu}\gamma_{AB}, \quad \mu \text{ and } \nu \text{ on different atoms, A and B} \tag{2.242}$$

Equation (2.240) is rather untidy, involving summations over basis functions on atom A and basis functions not on atom A. It is often simplified by writing P_{AA} as the total electron density on atom A, where:

$$P_{AA} = \sum_{\lambda \text{ on A}}^{A} P_{\lambda\lambda} \tag{2.243}$$

A similar expression can also be introduced for P_{BB}. With this notation $F_{\mu\mu}$ simplifies to:

$$F_{\mu\mu} = H_{\mu\mu}^{\text{core}} + (P_{AA} - \tfrac{1}{2}P_{\mu\mu})\gamma_{AA} + \sum_{B \neq A} P_{BB}\gamma_{AB} \tag{2.244}$$

The core Hamiltonian expressions, $H_{\mu\mu}^{\text{core}}$ and $H_{\mu\nu}^{\text{core}}$, correspond to electrons moving in the field of the parent nucleus and the other nuclei. In semi-empirical methods the core electrons are subsumed into the nucleus and so the nuclear charges are altered accordingly (for example, carbon has a nuclear 'charge' of +4).

In CNDO $H_{\mu\mu}^{\text{core}}$ is separated into an integral involving the atom on which ϕ_μ is situated (labelled A), and all the others (labelled B). Thus:

$$H_{\mu\mu}^{\text{core}} = U_{\mu\mu} - \sum_{B \neq A} V_{AB} \tag{2.245}$$

where:

$$U_{\mu\mu} = \left(\mu\left| -\frac{1}{2}\nabla^2 - \frac{Z_A}{|\mathbf{r}_1 - \mathbf{R}_A|}\right|\mu\right) \quad \text{and} \quad V_{AB} = \left(\mu\left|\frac{Z_B}{|\mathbf{r}_1 - \mathbf{R}_B|}\right|\mu\right) \tag{2.246}$$

$U_{\mu\mu}$ is thus the energy of the orbital ϕ_μ in the field of its own nucleus (A) and core electrons; $-V_{AB}$ is the energy of the electron in the field of another nucleus (B). To maintain consistency with the way in which the two-electron integrals are treated, the terms

$$\left(\mu\left|\frac{Z_B}{|\mathbf{r}_1 - \mathbf{R}_B|}\right|\mu\right) \tag{2.247}$$

must be the same for all orbitals ϕ_μ on atom A (i.e. the interaction energy between any electron in an orbital on atom A with the core of atom B is equal to V_{AB}).

We next consider $H_{\mu\nu}^{\text{core}}$, where ϕ_μ and ϕ_ν are both on the same atom, A. In this case the core Hamiltonian has the following form:

$$H_{\mu\nu}^{\text{core}} = \left(\mu\left| -\frac{1}{2}\nabla^2 - \frac{Z_A}{|\mathbf{r}_1 - \mathbf{R}_A|}\right|\nu\right) - \sum_{B\neq A}\left(\mu\left|\frac{Z_B}{|\mathbf{r}_1 - \mathbf{R}_B|}\right|\nu\right)$$

$$= U_{\mu\nu} - \sum_{B\neq A}\left(\mu\left|\frac{Z_B}{|\mathbf{r}_1 - \mathbf{R}_B|}\right|\nu\right) \tag{2.248}$$

As ϕ_μ and ϕ_ν are on the same atom, $U_{\mu\nu}$ is zero due to the orthogonality of atomic orbitals. The term

$$\left(\mu\left|\frac{Z_B}{|\mathbf{r}_1 - \mathbf{R}_B|}\right|\nu\right) \tag{2.249}$$

is zero in accordance with the zero-differential overlap approximation. Thus $H_{\mu\nu}^{\text{core}}$ is zero in CNDO.

Finally, if ϕ_μ and ϕ_ν are on two different atoms A and B, then we can write:

$$H_{\mu\nu}^{\text{core}} = \left(\mu\left| -\frac{1}{2}\nabla^2 - \frac{Z_A}{|\mathbf{r}_1 - \mathbf{R}_A|} - \frac{Z_B}{|\mathbf{r}_1 - \mathbf{R}_B|}\right|\nu\right) - \sum_{C\neq A,B}\left(\mu\left| -\frac{Z_C}{|\mathbf{r}_1 - \mathbf{R}_C|}\right|\nu\right) \tag{2.250}$$

The second term corresponds to the interaction of the distribution $\phi_\mu\phi_\nu$ with the atoms C ($\neq$ A, B). These interactions are ignored. The first part (known as the *resonance integral* and commonly written $\beta_{\mu\nu}$) is not subject to the ZDO approximation, because it is the main cause of bonding. In CNDO the resonance integral is made proportional to the overlap integral, $S_{\mu\nu}$:

$$H_{\mu\nu}^{\text{core}} = \beta_{AB}^0 S_{\mu\nu} \tag{2.251}$$

where β_{AB}^0 is a parameter which depends on the nature of atoms A and B.

With these approximations the Fock matrix elements for CNDO become:

$$F_{\mu\mu} = U_{\mu\mu} + \sum_{B\neq A} V_{AB} + (P_{AA} - \tfrac{1}{2}P_{\mu\mu})\gamma_{AA} + \sum_{B\neq A} P_{BB}\gamma_{AB} \tag{2.252}$$

$$F_{\mu\nu} = -\tfrac{1}{2}P_{\mu\nu}\gamma_{AA}; \quad \mu \text{ and } \nu \text{ on the same atom, A} \tag{2.253}$$

$$F_{\mu\nu} = \beta_{AB}^0 S_{\mu\nu} - \tfrac{1}{2}P_{\mu\nu}\gamma_{AB}, \quad \mu \text{ on A and } \nu \text{ on B} \tag{2.254}$$

To perform a CNDO calculation requires the following to be calculated or specified: the overlap integrals, $S_{\mu\nu}$, the core Hamiltonians $U_{\mu\mu}$, the electron–core interactions V_{AB}, the electron repulsion integrals γ_{AB} and γ_{AA} and the bonding parameters β_{AB}^0. The CNDO basis set comprises Slater type orbitals for the valence shell with the exponents being chosen using Slater's rules (except for hydrogen, where an exponent of 1.2 is used as this value is more appropriate to hydrogen atoms in molecules). Thus the basis set comprises 1s for hydrogen and 2s, $2p_x$, $2p_y$ and $2p_z$ for the first-row elements. The overlap integrals are calculated explicitly (the overlap between two basis functions on the same atom is, of course, zero with an s, p basis set). The electron repulsion integral parameter γ_{AB} is

calculated using valence s functions on the two atoms A and B:

$$\gamma_{AB} = \iint d\nu_1\, d\nu_2 \phi_{s,A}(1)\phi_{s,A}(1)\left(\frac{1}{r_{12}}\right)\phi_{s,B}(2)\phi_{s,B}(2) \tag{2.255}$$

The use of spherically symmetric s orbitals avoids the problems associated with transformations of the axes. The core Hamiltonians ($U_{\mu\mu}$) are not calculated but are obtained from experimental ionisation energies. This is because it is important to distinguish between s and p orbitals in the valence shell (i.e. the 2s and 2p orbitals for the first-row elements), and without explicit core electrons this is difficult to achieve. The resonance integrals, β^0_{AB}, are written in terms of empirical single-atom values as follows:

$$\beta^0_{AB} = \tfrac{1}{2}(\beta^0_A + \beta^0_B) \tag{2.256}$$

The β^0 values are chosen to fit the results of minimal basis set *ab initio* calculations on diatomic molecules.

The electron–core interaction, V_{AB}, is calculated as the interaction between an electron in a valence s orbital on atom A with the nuclear core of atom B:

$$V_{AB} = \int d\nu_1 \phi_{s,A}(1)\frac{Z_B}{|\mathbf{r}_1 - \mathbf{R}_B|}\phi_{s,A}(1) \tag{2.257}$$

CNDO is rightly recognised as the first in a long line of important semi-empirical models. However, there were some important limitations with the model. One especially serious deficiency of the first version of CNDO (introduced in 1965 [Pople and Segal 1965, Pople *et al.* 1965] and now known as CNDO/1) is that two neutral atoms show a significant (and incorrect) attraction, even when separated by several ångströms. The predicted equilibrium distances for diatomic molecules are also too short and the dissociation energies too large. These effects are due to electrons on one atom penetrating the valence shell of another atom and so experiencing a nuclear attraction. This penetration effect can be quantified more explicitly as follows. The net charge on an atom B equals the difference between its nuclear charge and the total electron density: $Q_B = Z_B - P_{BB}$. If we now substitute for P_{BB} ($= Z_B - Q_B$) in the diagonal elements of the Fock matrix, Equation (2.252), we obtain:

$$F_{\mu\mu} = U_{\mu\mu} + (P_{AA} - \tfrac{1}{2}P_{\mu\mu})\gamma_{AA} + \sum_{B\neq A}[-Q_B\gamma_{AB} + (Z_B\gamma_{AB} - V_{AB})] \tag{2.258}$$

$-Q_B\gamma_{AB}$ is the contribution from the total charge on atom B; this is zero if the atomic charge is exactly balanced by the electron density. $Z_B\gamma_{AB} - V_{AB}$ is called the *penetration integral*. It was this contribution that caused the anomalous results for two neutral atoms at large separation. In the second version of CNDO (CNDO/2 [Pople and Segal 1966]) the penetration integral effect was eliminated by putting $V_{AB} = Z_B\gamma_{AB}$. The core Hamiltonian $U_{\mu\mu}$ was also defined differently in CNDO/2, using both ionisation energies and electron affinities.

2.9.3 INDO

CNDO makes no allowance for the fact that the interaction between two electrons depends upon their relative spins. This effect can be particularly severe for electrons on the same

atom. Thus, in CNDO all two-electron integrals $(\mu\nu|\lambda\nu)$ are set to zero, and integrals $(\mu\mu|\nu\nu)$ and $(\mu\mu|\mu\mu)$ are forced to be equal (to γ_{AA}). The next development was the intermediate neglect of differential overlap model (INDO [Pople *et al.* 1967]), which includes monatomic differential overlap for one-centre integrals (i.e. for integrals involving basis functions centred on the same atom) This enables the interaction between two electrons on the same atom with parallel spins to have a lower energy than the comparable interaction between electrons with paired spins. For this reason the Fock matrix elements are usually written with the spin (α or β) explicitly specified. The elements $F_{\mu\mu}$ and $F_{\mu\nu}$ (where μ and ν are located on atom A) then change from their CNDO/2 values as follows:

$$F^{\alpha}_{\mu\mu} = U_{\mu\mu} + \sum_{\lambda \text{ on A}} \sum_{\sigma \text{ on A}} [P_{\lambda\sigma}(\mu\mu|\lambda\sigma) - P^{\alpha}_{\lambda\sigma}(\mu\lambda|\mu\sigma)] + \sum_{B \neq A} (P_{BB} - Z_B)\gamma_{AB} \tag{2.259}$$

$$F^{\alpha}_{\mu\nu} = U_{\mu\nu} + \sum_{\lambda \text{ on A}} \sum_{\sigma \text{ on A}} [P_{\lambda\sigma}(\mu\nu|\lambda\sigma) - P^{\alpha}_{\lambda\sigma}(\mu\lambda|\nu\sigma)]; \quad \mu \text{ and } \nu \text{ both on atom A} \tag{2.260}$$

In Equation (2.259) we have included the CNDO/2 approximation $V_{AB} = Z_B\gamma_{AB}$. The matrix element $F_{\mu\nu}$, where μ and ν are on different atoms, is the same as in CNDO/2:

$$F^{\alpha}_{\mu\nu} = \tfrac{1}{2}(\beta^0_A + \beta^0_B)S_{\mu\nu} - P^{\alpha}_{\mu\nu}\gamma_{AB} \tag{2.261}$$

In a closed-shell system, $P^{\alpha}_{\mu\nu} = P^{\beta}_{\mu\nu} = \frac{1}{2}P_{\mu\nu}$ and the Fock matrix elements can be obtained by making this substitution. If a basis set containing s, p orbitals is used, then many of the one-centre integrals nominally included in INDO are equal to zero, as are the core elements $U_{\mu\nu}$. Specifically, only the following one-centre, two-electron integrals are non-zero: $(\mu\mu|\mu\mu)$, $(\mu\mu|\nu\nu)$ and $(\mu\nu|\mu\nu)$. The elements of the Fock matrix that are affected can then be written as follows:

$$F_{\mu\mu} = U_{\mu\mu} + \sum_{\nu \text{ on A}} [P_{\nu\nu}(\mu\mu|\nu\nu) - \tfrac{1}{2}P_{\nu\nu}(\mu\nu|\mu\nu)] + \sum_{B \neq A} (P_{BB} - Z_B)\gamma_{AB} \tag{2.262}$$

$$F_{\mu\nu} = \tfrac{3}{2}P_{\mu\nu}(\mu\nu|\mu\nu) - \tfrac{1}{2}P_{\mu\nu}(\mu\mu|\nu\nu); \quad \mu, \nu \text{ on the same atom} \tag{2.263}$$

Some of the one-centre two-electron integrals in INDO are semi-empirical parameters, obtained by fitting to atomic spectroscopic data. The core integrals $U_{\mu\mu}$ are obtained in a slightly different fashion to that of CNDO/2, to take into account the new electronic configurations under the INDO model for atoms and their cations and anions. An INDO calculation requires little additional computational effort compared with the corresponding CNDO calculation and has the key advantage that states of different multiplicities can be distinguished. For example, in CNDO the singlet and triplet configurations $1s^2 2s^2 2p^2$ of carbon have the same energy, whereas these can be distinguished using INDO. Two of the systems considered in the original INDO publication were the methyl and ethyl radicals, the unpaired electron density being compared with experimentally determined hyperfine coupling constants. INDO gave a much more favourable result for these systems than CNDO.

2.9.4 NDDO

The next level of approximation is the neglect of diatomic differential overlap model (NDDO [Pople *et al.* 1965]); this theory only neglects differential overlap between atomic orbitals on

different atoms. Thus all of the two-electron, two-centre integrals of the form $(\mu\nu|\lambda\sigma)$, where μ and ν are on the same atom and λ and σ are also on the same atom, are retained. The Fock matrix elements become:

$$F_{\mu\mu} = H_{\mu\mu}^{\text{core}} + \sum_{\lambda \text{ on A}} \sum_{\sigma \text{ on A}} [P_{\lambda\sigma}(\mu\mu|\lambda\sigma) - \tfrac{1}{2}P_{\lambda\sigma}(\mu\lambda|\mu\sigma)] + \sum_{\text{B}\neq\text{A}} \sum_{\lambda \text{ on B}} \sum_{\sigma \text{ on B}} P_{\lambda\sigma}(\mu\mu|\lambda\sigma) \quad (2.264)$$

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + \sum_{\lambda \text{ on A}} \sum_{\sigma \text{ on A}} [P_{\lambda\sigma}(\mu\nu|\lambda\sigma) - \tfrac{1}{2}P_{\lambda\sigma}(\mu\lambda|\nu\sigma)] + \sum_{\text{B}\neq\text{A}} \sum_{\lambda \text{ on B}} \sum_{\sigma \text{ on B}} P_{\lambda\sigma}(\mu\nu|\lambda\sigma); \quad \mu \text{ and } \nu \text{ both on A} \quad (2.265)$$

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} - \frac{1}{2} \sum_{\lambda \text{ on B}} \sum_{\sigma \text{ on A}} P_{\lambda\sigma}(\mu\sigma|\nu\lambda); \quad \mu \text{ on A and } \nu \text{ on B} \quad (2.266)$$

It is again possible to tidy up equations (2.264) and (2.265) when an s, p basis set is used:

$$F_{\mu\mu} = H_{\mu\mu}^{\text{core}} + \sum_{\nu \text{ on A}} [P_{\nu\nu}(\mu\mu|\nu\nu) - \tfrac{1}{2}P_{\nu\nu}(\mu\nu|\mu\nu)] + \sum_{\text{B}\neq\text{A}} \sum_{\lambda \text{ on B}} \sum_{\sigma \text{ on B}} P_{\lambda\sigma}(\mu\mu|\lambda\sigma) \quad (2.267)$$

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + \tfrac{3}{2}P_{\mu\nu}(\mu\nu|\mu\nu) - \tfrac{1}{2}P_{\mu\nu}(\mu\mu|\nu\nu) + \sum_{\text{B}\neq\text{A}} \sum_{\lambda \text{ on B}} \sum_{\sigma \text{ on B}} P_{\lambda\sigma}(\mu\nu|\lambda\sigma) \quad (2.268)$$

Whereas the computation required for an INDO calculation is little more than for the analogous CNDO calculation, in NDDO the number of two-electron, two-centre integrals is increased by a factor of approximately 100 for each pair of heavy atoms in the system.

2.9.5 MINDO/3

The CNDO, INDO and NDDO methods, as originally devised and implemented, are now little used, in comparison with the methods subsequently developed by Dewar and colleagues, but they were of considerable importance in showing how a systematic series of approximations could be used to develop methods of real practical value. Moreover, the calculations could be performed in a fraction of the time required to solve the full Roothaan–Hall equations. However, they did not produce very accurate results, largely because they were parametrised upon the results from relatively low-level *ab initio* calculations, which themselves agreed poorly with experiment. They were also limited to small classes of molecule, and they often required a good experimental geometry to be supplied as input because their geometry optimisation algorithms were not very sophisticated.

It was through the introduction of the MINDO/3 method by Bingham, Dewar and Lo [Bingham *et al.* 1975a–d] that a wider audience was able to apply semi-empirical methods in their own research. MINDO/3 was not so much a significant change in the theory, being based upon INDO (MINDO stands for modified INDO), but it did differ significantly in the way in which the method was parametrised, making much more use of experimental data. It also incorporated a geometry optimisation routine (the Davidon–Fletcher–Powell method; see Chapter 5), which enabled the program to accept crude initial geometries as input and derive the associated minimum energy structures.

MINDO/3 uses an s, p basis set and its Fock matrix elements are:

$$F_{\mu\mu} = U_{\mu\mu} + \sum_{\nu \text{ on A}} (P_{\nu\nu}(\mu\mu|\nu\nu) - \tfrac{1}{2}P_{\nu\nu}(\mu\nu|\mu\nu)) + \sum_{B \neq A} (P_{BB} - Z_B)\gamma_{AB} \tag{2.269}$$

$$F_{\mu\nu} = -\tfrac{1}{2}P_{\mu\nu}(\mu\nu|\mu\nu); \quad \mu \text{ and } \nu \text{ both on the same atom A} \tag{2.270}$$

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} - \tfrac{1}{2}P_{\mu\nu}(\mu\nu|\mu\nu) = H_{\mu\nu}^{\text{core}} - \tfrac{1}{2}P_{\mu\nu}\gamma_{AB}; \quad \mu \text{ on A and } \nu \text{ on B} \tag{2.271}$$

The two-centre repulsion integrals γ_{AB} in MINDO/3 are calculated using the following function.

$$\gamma_{AB} = \frac{e^2}{\left[R_{AB}^2 + \frac{1}{4}\left(\frac{e^2}{\bar{g}_A} + \frac{e^2}{\bar{g}_B}\right)^2\right]^{1/2}} \tag{2.272}$$

$\bar{g}_A$ is the average of the one-centre, two-electron integrals $g_{\mu\nu}$ on atom A (i.e. $g_{\mu\nu} \equiv (\mu\mu|\nu\nu)$) and $\bar{g}_B$ is the equivalent average for atom B. This seemingly complex function for γ_{AB} is, in fact, quite simple; at large R_{AB} it tends towards the Coulomb's law expression e^2/R_{AB} and as R_{AB} tends to zero it approaches the average of the one-centre integrals on the two atoms. The two-centre, one-electron integrals $H_{\mu\nu}^{\text{core}}$ are given in MINDO/3 by:

$$H_{\mu\nu}^{\text{core}} = S_{\mu\nu}\beta_{AB}(I_\nu + I_\nu) \tag{2 273}$$

$S_{\mu\nu}$ is the overlap integral, I_μ and I_ν are ionisation potentials for the appropriate orbitals and β_{AB} is a parameter dependent upon both of the two atoms A and B.

The core–core interaction between pairs of nuclei was also changed in MINDO/3 from the form used in CNDO/2. One way to correct the fundamental problems with CNDO/2 such as the repulsion between two hydrogen atoms (or indeed any neutral molecules) at all distances is to change the core–core repulsion term from a simple Coulombic expression ($E_{AB} = Z_A Z_B/R_{AB}$) to:

$$E_{AB} = Z_A Z_B \gamma_{AB} \tag{2.274}$$

In fact, while this correction gives the desired behaviour at relatively long separations, it does not account for the fact that as two nuclei approach each other the screening by the core electrons decreases. As the separation approaches zero the core–core repulsion should be described by Coulomb's law. In MINDO/3 this is achieved by making the core–core interaction a function of the electron–electron repulsion integrals as follows:

$$E_{AB} = Z_A Z_B\{\gamma_{AB} + [(e^2/R_{AB}) - \gamma_{AB}]\exp(-\alpha_{AB}R_{AB})\} \tag{2.275}$$

α_{AB} is a parameter dependent upon the nature of the atoms A and B. For OH and NH bonds a slightly different core–core interaction was found to be more appropriate:

$$E_{XH} = Z_X Z_H\{\gamma_{XH} + [(e^2/R_{XH}) - \gamma_{XH}]\alpha_{XH}\exp(-R_{XH})\} \tag{2.276}$$

The parameters for MINDO/3 were obtained in an entirely different way from previous semi-empirical methods. Some of the values that were fixed in CNDO, INDO and NNDO were permitted to vary during the MINDO/3 parametrisation procedure. For example, the exponents of the Slater atomic orbitals were allowed to vary from the values given by Slater's rules, and indeed the exponents for s and p orbitals were not required to be the

same. $U_{\mu\mu}$ and β_{AB} were also regarded as variable parameters. Another key difference was that the MINDO/3 parametrisation used experimental data such as molecular geometries and heats of formation, rather than theoretical values from *ab initio* calculations or data from atomic spectra. The parametrisation effort was a considerable undertaking, and it was only at the fourth attempt that an acceptable model was obtained (as is implicit in the appearance of the '3' in the name). For example, just to parametrise two atoms such as carbon and hydrogen using a set of 20 molecules required between 30 000 and 50 000 SCF calculations for each parametrisation scheme that was investigated.

2.9.6 MNDO

MINDO/3 proved to be very successful when it was introduced; it is important to realise that even simple *ab initio* calculations were beyond the computational resources of all but a few research groups in the 1970s. However, there were some significant limitations. For example, heats of formation of unsaturated molecules were consistently too positive, the errors in calculated bond angles were often quite large, and the heats of formation for molecules containing adjacent atoms with lone pairs were too negative. Some of these limitations were due to the use of the INDO approximation, and in particular the inability of INDO to deal with systems containing lone pairs. Dewar and Thiel therefore introduced the modified neglect of diatomic overlap (MNDO) method, which was based on NDDO [Dewar and Thiel 1977a, b]. The Fock matrix elements in MNDO were as follows:

$$F_{\mu\mu} = H_{\mu\mu}^{\text{core}} + \sum_{\nu \text{ on A}} [P_{\nu\nu}(\mu\mu|\nu\nu) - \tfrac{1}{2}P_{\nu\nu}(\mu\nu|\mu\nu)] + \sum_{B\neq A}\sum_{\lambda \text{ on B}}\sum_{\sigma \text{ on B}} P_{\lambda\sigma}(\mu\mu|\lambda\sigma) \tag{2.277}$$

$$\text{where} \quad H_{\mu\mu}^{\text{core}} = U_{\mu\mu} - \sum_{B\neq A} V_{\mu\mu B} \tag{2.278}$$

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + \tfrac{3}{2}P_{\mu\nu}(\mu\nu|\mu\nu) - \tfrac{1}{2}P_{\mu\nu}(\mu\mu|\nu\nu) + \sum_{B\neq A}\sum_{\lambda \text{ on B}}\sum_{\sigma \text{ on B}} P_{\lambda\sigma}(\mu\nu|\lambda\sigma); \quad \mu \text{ and } \nu \text{ both on A} \tag{2.279}$$

$$\text{where} \quad H_{\mu\nu}^{\text{core}} = -\sum_{B\neq A} V_{\mu\nu B} \tag{2.280}$$

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} - \tfrac{1}{2}\sum_{\lambda \text{ on B}}\sum_{\sigma \text{ on A}} P_{\lambda\sigma}(\mu\sigma|\nu\lambda); \quad \mu \text{ on A and } \nu \text{ on B} \tag{2.281}$$

$$\text{where} \quad H_{\mu\nu}^{\text{core}} = \tfrac{1}{2}S_{\mu\nu}(\beta_\mu + \beta_\nu) \tag{2.282}$$

The similarity with the NDDO expressions, Equations (2.264)–(2.266), can clearly be seen; the major new features are the appearance of terms $V_{\mu\mu B}$ and $V_{\mu\nu B}$ and a new form for the two-centre, one-electron core resonance integrals, which depend upon the overlap $S_{\mu\nu}$ and parameters β_μ and β_ν as shown in Equation (2.282). $V_{\mu\mu B}$ and $V_{\mu\nu B}$ are two-centre, one-electron attractions between an electron distribution $\phi_\mu\phi_\mu$ or $\phi_\mu\phi_\nu$, respectively, on atom A and the core of atom B. These are expressed as follows:

$$V_{\mu\mu B} = -Z_B(\mu_A\mu_A|s_B s_B) \tag{2.283}$$

$$V_{\mu\nu B} = -Z_B(\mu_A\nu_A|s_B s_B) \tag{2.284}$$

The core–core repulsion terms are also different in MNDO from those in MINDO/3, with OH and NH bonds again being treated separately:

$$E_{AB} = Z_A Z_B (s_A s_A | s_B s_B)\{1 + \exp(-\alpha_A R_{AB}) + \exp(-\alpha_B R_{AB})\} \quad (2.285)$$

$$E_{XH} = Z_X Z_H (s_X s_X | s_H s_H)\{1 + R_{XH}\exp(-\alpha_X R_{XH})/R_{AB}) + \exp(-\alpha_H R_{XH})\} \quad (2.286)$$

Perhaps the most significant advantage of MNDO over MINDO/3 is the use throughout of monatomic parameters; MINDO/3 requires diatomic parameters in the resonance integral (β_{AB}) and the core–core repulsion (α_{AB}). It has been possible to expand MNDO to cover a much wider variety of elements such as aluminium, silicon, germanium, tin, bromine and lead. However, the use of an (s, p) basis set in the original MNDO method did mean that the method could not be applied to most transition metals, which require a basis set containing d orbitals. In addition, hypervalent compounds of sulphur and phosphorus are not modelled well. In more recent versions of the MNDO method d orbitals have been explicitly included for the heavier elements [Thiel and Voityuk 1994]. Another serious limitation of MNDO is its inability to accurately model intermolecular systems involving hydrogen bonds (for example, the heat of formation of the water dimer is far too low in MNDO). This is because of a tendency to overestimate the repulsion between atoms when they are separated by a distance approximately equal to the sum of their van der Waals radii. Conjugated systems can also present difficulties for MNDO. An extreme example of this occurs with compounds such as nitrobenzene in which the nitro group is predicted to be orthogonal to the aromatic ring rather than conjugated with it. In addition, MNDO energies are too positive for sterically crowded molecules and too negative for molecules containing four-membered rings.

2.9.7 AM1

The Austin Model 1 (AM1) model was the next semi-empirical theory produced by Dewar's group [Dewar *et al.* 1985]. AM1 was designed to eliminate the problems with MNDO, which were considered to arise from a tendency to overestimate repulsions between atoms separated by distances approximately equal to the sum of their van der Waals radii. The strategy adopted was to modify the core–core term using Gaussian functions. Both attractive and repulsive Gaussian functions were used; the attractive Gaussians were designed to overcome the repulsion directly and were centred in the region where the repulsions were too large Repulsive Gaussian functions were centred at smaller internuclear separations. With this modification the expression for the core–core term was related to the MNDO expression by:

$$E_{AB} = E_{MNDO} + \frac{Z_A Z_B}{R_{AB}} \times \left\{ \sum_i K_{A_i} \exp[-L_{A_i}(R_{AB} - M_{A_i})^2] + \sum_j K_{B_j} \exp[-L_{B_j}(R_{AB} - M_{B_j})^2] \right\} \quad (2.287)$$

The additional terms are spherical Gaussian functions with a width determined by the parameter L. It was found that the values of these parameters were not critical and many

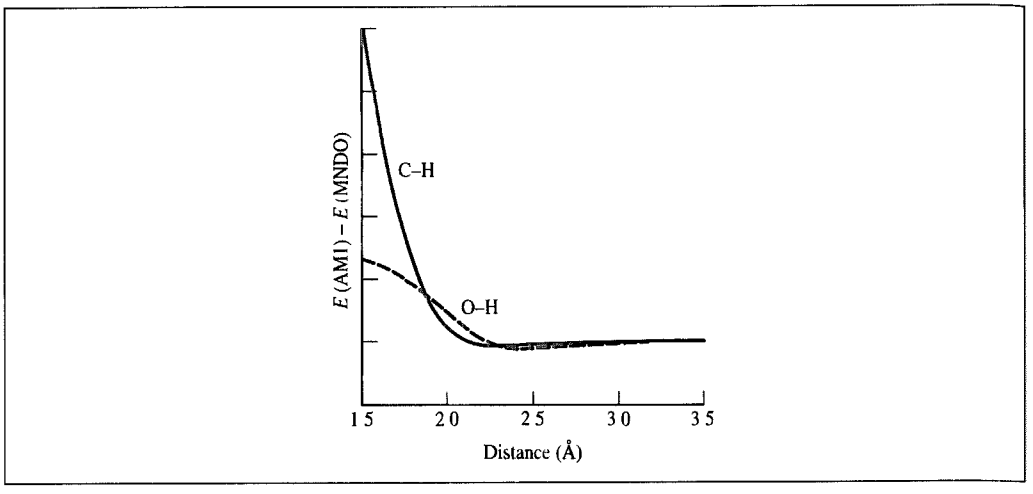

Fig 2 20. The difference in the core-core energy for AM1 and MNDO for carbon-hydrogen and oxygen-hydrogen interactions

were set to the same value. The M and K parameters were optimised for each atom, together with the α parameters in the exponential terms in Equations (2.285) and (2.286). In the original parametrisation of AM1 there are four terms in the Gaussian expansion for carbon, three for hydrogen and nitrogen and two for oxygen (both attractive and repulsive Gaussians were used for carbon, hydrogen and nitrogen but only repulsive Gaussians for oxygen). The effect of including these Gaussian functions can be seen in Figure 2.20, which plots the difference in the MNDO and AM1 core-core terms for the carbon-hydrogen and oxygen-hydrogen interactions. The inclusion of these Gaussians significantly increased the number of parameters per atom, from seven in the MNDO to between 13 and 16 per atom in AM1. This, of course, made the parametrisation process considerably more difficult. Overall, AM1 was a significant improvement over MNDO and many of the deficiencies associated with the core repulsion were corrected.

2.9.8 PM3

PM3 is also based on MNDO (the name derives from the fact that it is the third parametrisation of MNDO, AM1 being considered the second) [Stewart 1989a, b]. The PM3 Hamiltonian contains essentially the same elements as that for AM1, but the parameters for the PM3 model were derived using an automated parametrisation procedure devised by J J P Stewart By contrast, many of the parameters in AM1 were obtained by applying chemical knowledge and 'intuition'. As a consequence, some of the parameters have significantly different values in AM1 and PM3, even though both methods use the same functional form and they both predict various thermodynamic and structural properties to approximately the same level of accuracy. Some problems do remain with PM3. One of the most important of these is the rotational barrier of the amide bond, which is much too low and in some cases almost non-existent. This problem can be corrected through the use of an

empirical torsional potential (see Section 4.5). There has been considerable debate over the relative merits of the AM1 and PM3 approaches to parametrisation.

2.9.9 SAM1

The final offering from the Dewar group* was SAM1, which stands for 'Semi-Ab-initio Model 1' [Dewar *et al.* 1993]. The name was chosen to reflect Dewar's belief that methods like AM1 offer such a significant enhancement over the earlier semi-empirical methods like CNDO/2 that they should be given a different generic name. In SAM1 a standard STO-3G Gaussian basis set is used to evaluate the electron repulsion integrals; close inspection of the results from AM1 and MNDO suggested that steric effects were overestimated because of the way in which the electron repulsion integrals were calculated. The resulting integrals were then scaled, partly to enable some of the effects of electron correlation to be included and partly to compensate for the use of a minimal basis set. The Gaussian terms in the core–core repulsion were retained to fine-tune the model. The number of parameters in SAM1 is no greater than in AM1 and fewer than in PM3. It does take longer to run (by up to two orders of magnitude) though it was felt that with the improvements in computer hardware such an increase was acceptable.

2.9.10 Programs for Semi-empirical Quantum Mechanical Calculations

The popularity of the MNDO, AM1 and PM3 methods is due in large part to their implementation in the MOPAC and AMPAC programs. The programs are able to perform many kinds of calculation and to calculate many different properties.

The contributions of the Dewar group are rightly recognised as particularly significant in the development of semi-empirical methods, but other research groups have also made important contributions. The SINDO1 and ZINDO programs have been developed in the groups of Jug and Zerner, respectively, and both contain novel features. The ZINDO program of Zerner and co-workers can perform a wide variety of semi-empirical calculations and has been particularly useful for calculations on transition metal and lanthanide compounds and for predicting molecular electronic spectra.

2.10 Hückel Theory

Hückel theory can be considered the 'grandfather' of approximate molecular orbital methods, having been formulated in the early 1930s [Hückel 1931]. Hückel theory is limited to conjugated π systems and was originally devised to explain the non-additive nature of certain properties of aromatic compounds. For example, the properties of benzene are much different from those of the hypothetical 'cyclohexatriene' molecule. Although Hückel theory, as originally formulated, is relatively little used in research today, extensions

*Michael Dewar died in 1997

to it such as extended Hückel theory are still employed and can provide qualitative insights into the electronic structure of important classes of molecule. Hückel theory is also widely used for teaching purposes to introduce a 'real' theory that can be applied to relatively complex systems with little more than pencil and paper or a simple computer program.

Hückel theory separates the π system from the underlying σ framework and constructs molecular orbitals into which the π electrons are then fed in the usual way according to the Aufbau principle. The π electrons are thus considered to be moving in a field created by the nuclei and the 'core' of σ electrons. The molecular orbitals are constructed from linear combinations of atomic orbitals and so the theory is an LCAO method. For our purposes it is most appropriate to consider Hückel theory in terms of the CNDO approximation (in fact, Hückel theory was the first ZDO molecular orbital theory to be developed). Let us examine the three types of Fock matrix element in Equations (2.252)–(2.254). First, $F_{\mu\mu}$. In a neutral species, the net charge on each atom will be approximately zero, and so if we take Equation (2.258), from which penetration effects have been eliminated, then we are left with $U_{\mu\mu} + (P_{AA} - 0\,5P_{\mu\mu})\gamma_{AA}$. Now if each nucleus (A) in the π system is the same (i.e. carbon) then this expression will be approximately constant for all nuclei being considered. The matrix elements $F_{\mu\mu}$ are often (confusingly) called Coulomb integrals in Hückel theory and are assigned the symbol α. All off-diagonal elements of the Fock matrix are assumed to be zero with the exception of elements $F_{\mu\nu}$, where μ and ν are π orbitals on two bonded atoms. These $F_{\mu\nu}$ are assumed to be constant, are assigned the symbol β and are known as resonance integrals. The Fock matrix in Hückel theory thus has as many rows and columns as the number of atoms in the π system with diagonal elements that are all set to α. All off-diagonal elements F_{ij} are zero unless there is a bond between the atoms i and j, in which case the element is β. For benzene the Fock matrix is of the following form (atom labelling as in Figure 2.21):

$$\begin{pmatrix} \alpha & \beta & 0 & 0 & 0 & \beta \\ \beta & \alpha & \beta & 0 & 0 & 0 \\ 0 & \beta & \alpha & \beta & 0 & 0 \\ 0 & 0 & \beta & \alpha & \beta & 0 \\ 0 & 0 & 0 & \beta & \alpha & \beta \\ \beta & 0 & 0 & 0 & \beta & \alpha \end{pmatrix} \tag{2.288}$$

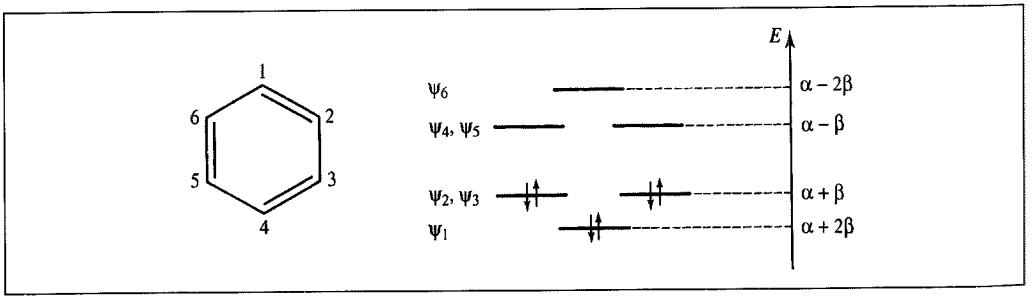

Fig 2 21: Benzene and its Hückel molecular orbitals.

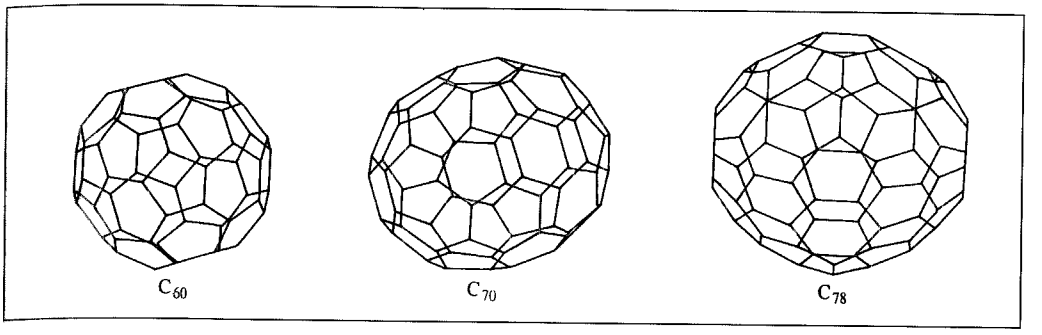

Fig 2 22 Three fullerenes, C_{60}, C_{70} and C_{78}.

As with the other semi-empirical methods that we have considered so far, the overlap matrix is equal to the identity matrix. The following simple matrix equation must then be solved:

$$\mathbf{FC} = \mathbf{CE} \tag{2 289}$$

The equation can be solved by standard methods to give the basis set coefficients and the molecular orbital energies E. The orbital energies for benzene are $E_1 = \alpha + 2\beta$; E_2, $E_3 = \alpha + \beta$; E_4, $E_5 = \alpha - \beta$; $E_6 = \alpha - 2\beta$, and so the ground state places two electrons in ψ_1 and two each in the two degenerate orbitals ψ_2 and ψ_3. The lowest-energy orbital ψ_1 is a linear combination of the six carbon p orbitals.

Hückel theory was extended to cover various other systems, including those with hetero-atoms, but it was not particularly successful and has largely been superseded by other semi-empirical methods. Nevertheless, for appropriate problems Hückel theory can be very useful. One example is the calculations of P W Fowler and colleagues, who studied the relationship between geometry and electronic structure for a range of buckminster-fullerenes (the parent molecule of which, C_{60}, was discovered in 1985) [Fowler 1993] The fullerenes (or 'buckyballs') are excellent candidates for Hückel theory as they are composed of carbon and have extensive π systems; three examples are shown in Figure 2.22.

The results of their calculations were summarised in two rules. The first rule states that at least one isomer C_n with a properly closed p shell (i.e. bonding HOMO, antibonding LUMO) exists for all $n = 60 + 6k$ ($k = 0, 2, 3, \ldots$, but not 1). Thus C_{60}, C_{72}, C_{78}, etc., are in this group. The second rule is for carbon cylinders and states that a closed-shell structure is found for $n = 2p(7 + 3k)$ (for all k). C_{70} is the parent of this family. The calculations were extended to cover different types of structure and fullerenes doped with metals.

2.10.1 Extended Hückel Theory

Hückel theory is clearly limited, in part because it is restricted to π systems. The extended Hückel method is a molecular orbital theory that takes account of all the valence electrons in the molecule [Hoffmann 1963]. It is largely associated with R Hoffmann, who received the Nobel Prize for his contributions The equation to be solved is $\mathbf{FC} = \mathbf{SCE}$, with the

Fock matrix elements taking the following simple forms:

$$F_{\mu\mu}^{AA} = H_{\mu\mu} = -I_{\mu} \tag{2.290}$$

$$F_{\mu\nu}^{AB} = H_{\mu\nu} = -\tfrac{1}{2}K(I_{\mu} + I_{\nu})S_{\mu\nu} \tag{2.291}$$

In these equations, μ and ν are two atomic orbitals (e.g. Slater type orbitals), I_{μ} is the ionisation potential of the orbital and K is a constant, which was originally set to 1.75. The formula for the off-diagonal elements $H_{\mu}\nu$ (where μ and ν are on different atoms) was originally suggested by R S Mulliken. These off-diagonal matrix elements are calculated between all pairs of valence orbitals and so extended Hückel theory is not limited to π systems.

The extended Hückel approach has proved to be rather successful for such a simple theory; for example, the famous Woodward–Hoffmann rules (see Section 5.9.4) were based upon calculations using this model. Extended Hückel theory has found particular application in those areas where alternative theories cannot be used. This is largely due to the fact that the basis set requires no more than experimentally determined ionisation potentials. It is particularly useful for studying systems containing metals; these systems are problematic for many other methods due to the lack of suitable basis sets.

2.11 Performance of Semi-empirical Methods

Our discussion of the application of quantum mechanics calculations was not explicitly directed towards any particular quantum mechanical theory but was - implicitly at least - written with *ab initio* methods in mind. All of the properties we considered in Section 2.7 can also be determined using semi-empirical methods. Extensive tables detailing the performance of the popular semi-empirical methods have been published, both in the original papers and in review articles, some of which are listed at the end of this chapter. The parametrisation of the semi-empirical approaches typically includes geometrical variables, dipole moments, ionisation energies and heats of formation. In Table 2.7 we provide a summary of the performance of the MINDO/3, MNDO, AM1, PM3 and SAM1 semi-empirical methods from data supplied in the original publications. The performance of successive semi-empirical methods has gradually improved from one method to another, though one should always remember that anomalous results may be obtained for certain types of system. Some of these limitations were outlined in the discussion of the various semi-empirical methods. It is worth emphasising that some of the major drawbacks with the semi-empirical methods arise simply because one is trying to calculate properties that were not given a major consideration in the parametrisation process. For example, many of the molecules used for the parametrisation of the MNDO, AM1 and PM3 methods had little or no conformational flexibility and it is therefore not so surprising that some rotational barriers are not calculated with the same accuracy as (say) heats of formation. In addition, to achieve optimal performance for specific classes of molecules (e.g. the amino acids) or specific properties (e.g. conformational barriers) then it would be appropriate to include representative systems during the parametrisation procedure.

	MINDO/3	MNDO	AM1	PM3	SAM1	Reference
138 heats of formation (kcal/mol)	11 0	6.3				[Dewar and Thiel 1977b]
228 bond lengths	0.022 Å	0 014 Å				
91 angles	5.6°	2.8°				
57 dipole moments	0.49 D	0.38 D				
58 heats of formation of hydrocarbons (kcal/mol)	9.7	5.87	5.07			[Dewar *et al.* 1985]
80 heats of formation for species with N and/or O (kcal/mol)	11.69	6.64	5 88			
46 dipole moments	0.54 D	0.32 D	0.26 D			
29 ionisation energies	0.31 eV	0.39 eV	0.29 eV			
406 heats of formation (kcal/mol)			8 82	7.12	5.21	[Dewar *et al* 1993]
196 dipole moments			0.35 D	0.40 D	0.32 D	

Table 2.7 Comparison of quantities calculated with various semi-empirical methods.

Appendix 2.1 Some Common Acronyms Used in Computational Quantum Chemistry

AM1	Austin Model 1
AO	Atomic obital
B3LYP	Scheme for hybrid Hartree–Fock/density functional theory introduced by Becke
BLYP	Becke–Lee–Yang–Parr gradient-corrected functional for use with density functional theory
BSSE	Basis set superposition error
CASSCF	Complete active space self-consistent field
CI	Configuration interaction
CIS	Configuration interaction singles
CISD	Configuration interaction singles and doubles
CNDO	Complete neglect of differential overlap
DFT	Density functional theory
DIIS	Direct inversion of iterative subspace
DVP	Double zeta with polarisation
DZ	Double zeta
EHT	Extended Hückel theory
GVB	Generalised valence bond model
HF	Hartree–Fock
HOMO	Highest occupied molecular orbital
INDO	Intermediate neglect of differential overlap
LCAO	Linear combination of atomic orbitals
LDA	Local density approximation
LSDFT	Local spin density functional theory
LUMO	Lowest unoccupied molecular orbital
MBPT	Many-body perturbation theory
MINDO/3	Modified INDO version 3
MNDO	Modified neglect of diatomic overlap
MO	Molecular orbital
MP	Møller–Plesset
MP2, MP3, etc.	Møller–Plesset theory at second order, third order, etc.
NDDO	Neglect of diatomic differential overlap
PM3	Parametrisation 3 of MNDO
QCISD	Quadratic configuration interaction singles and doubles
QCISD(T)	Configuration interation method involving single, double and quadratic excitations with an estimated triple excitation
RHF	Restricted Hartree–Fock
SAM1	Semi-*Ab initio* Model 1
SCF	Self-consistent field
STO	Slater type orbital
STO-3G, STO-4G, etc.	Minimal basis sets in which 3, 4 etc, Gaussian functions are used to represent the atomic orbitals on an atom

UHF	Unrestricted Hartree-Fock
WVN	Correlation functional due to Wilk, Vosko and Nusair
ZDO	Zero differential overlap

Further Reading

Atkins P W 1991 *Quanta. A Handbook of Concepts*. Oxford, Oxford University Press

Atkins P W 1998. *Physical Chemistry*. 6th Edition. Oxford, Oxford University Press.

Atkins P W and R S Friedman 1996 *Molecular Quantum Mechanics*. Oxford, Oxford University Press

Clark T 1985 *A Handbook of Computational Chemistry: A Practical Guide to Chemical Structure and Energy Calculations*. New York, Wiley-Interscience

Dewar M J S 1969. *The Molecular Orbital Theory of Organic Chemistry*. New York, McGraw-Hill.

Hinchliffe A 1988 *Computational Quantum Chemistry* Chichester, John Wiley & Sons.

Hinchliffe A 1995. *Modelling Molecular Structures* Chichester, John Wiley & Sons.

Hirst D M 1990 *A Computational Approach to Chemistry*. Oxford, Blackwell Scientific.

Pople J A and D L Beveridge 1970. *Approximate Molecular Orbital Theory*. New York, McGraw-Hill

Richards W G and D L Cooper 1983. *Ab initio Molecular Orbital Calculations for Chemists*. 2nd Edition Oxford, Clarendon Press.

Schaeffer H F III (Editor) 1977. *Applications of Electronic Structure Theory* New York, Plenum Press

Schaeffer H F III (Editor) 1977. *Methods of Electronic Structure Theory*. New York, Plenum Press

Stewart J J P 1990. MOPAC. A Semi-Empirical Molecular Orbital Program. *Journal of Computer-Aided Molecular Design* 4:1–45

Stewart J J P 1990. Semi-empirical Molecular Orbital Methods. In Lipkowitz K B and D B Boyd (Editors). *Reviews in Computational Chemistry* Volume 1. New York, VCH Publishers, pp 45–82

Szabo A and N S Ostlund 1982. *Modern Quantum Chemistry Introduction to Advanced Electronic Structure Theory*. New York, McGraw-Hill.

Zerner M C 1991. Semi-empirical Molecular Orbital Methods. In Lipkowitz K B and D B Boyd (Editors). *Reviews in Computational Chemistry* Volume 2. New York, VCH Publishers, pp 313–366

References

Allinger N L, R S Grev, B F Yates and H F Schaeffer III 1990. The Syn Rotational Barrier in Butane *Journal of the American Chemical Society* **112**:114–118.

Bachrach S M 1994. Population Analysis and Electron Densities from Quantum Mechanics. In Lipkowitz K B and D B Boyd (Editors). *Reviews in Computational Chemistry* Volume 5. New York, VCH Publishers, pp 171–227.

Bader R F W 1985 Atoms in Molecules. *Accounts of Chemistry Research* **18**:9–15.

Bingham R C, M J S Dewar and D H Lo 1975a. Ground States of Molecules. XXV. MINDO/3. An Improved Version of the MINDO Semi-empirical SCFMO Method. *Journal of the American Chemical Society* **97**.1285–1293.

Bingham R C, M J S Dewar and D H Lo 1975b Ground States of Molecules XXVI. MINDO/3. Calculations for Hydrocarbons *Journal of the American Chemical Society* **97**:1294–1301.

Bingham R C, M J S Dewar and D H Lo 1975c. Ground States of Molecules XXVII MINDO/3 Calculations for CHON Species. *Journal of the American Chemical Society* **97**:1302–1306.

Bingham R C, M J S Dewar and D H Lo 1975d. Ground States of Molecules. XXVIII MINDO/3. Calculations for Compounds Containing Carbon, Hydrogen, Fluorine and Chlorine. *Journal of the American Chemical Society* **97** 1307–1310.

Boys S F 1950. Electronic Wave Functions. I. A General Method of Calculation for the Stationary States of Any Molecular System. *Proceedings of the Royal Society (London)* **A200**:542–554.

Cusachs L C and Politzer 1968. On the Problem of Defining the Charge on an Atom in a Molecule *Chemical Physics Letters* **1**:529–531.

Dewar M J S, C Jie and J Yu 1993. SAM1; The First of a New Series of General Purpose Quantum Mechanical Molecular Models. *Tetrahedron* **49**:5003–5038.

Dewar M J S and Thiel W 1977a. Ground States of Molecules. 38. The MNDO Method Approximations and Parameters. *Journal of the American Chemical Society* **99**:4899–4907.

Dewar M J S and Thiel W 1977b. Ground States of Molecules 39. MNDO Results for Molecules Containing Hydrogen, Carbon, Nitrogen and Oxygen. *Journal of the American Chemical Society* **99**:4907–4917

Dewar M J S, E G Zoebisch, E F Healy and J J P Stewart 1985. AM1: A New General Purpose Quantum Mechanical Model. *Journal of the American Chemical Society* **107**:3902–3909

Dunning T H Jr 1970 Gaussian Basis Functions for Use in Molecular Calculations. I. Contraction of (9s5p) Atomic Basis Sets for First-Row Atoms. *Journal of Chemical Physics* **53**:2823–2883.

Fowler P W 1993 Systematics of Fullerenes and Related Clusters *Philosophical Transactions of the Royal Society (London)* **A343**:39–52.

Hall G G 1951 The Molecular Orbital Theory of Chemical Valency VIII A Method for Calculating Ionisation Potentials. *Proceedings of the Royal Society (London)* **A205**:541–552

Hehre W J, R F Stewart and J A Pople 1969 Self-Consistent Molecular-Orbital Methods I Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *Journal of Chemical Physics* **51**:2657–2664

Hehre W J, L Radom, P v R Schleyer and J A Pople 1986 *Ab initio Molecular Orbital Theory* New York, John Wiley & Sons.

Hoffmann R 1963. An Extended Hückel Theory. I. Hydrocarbons. *Journal of Chemical Physics* **39**:1397–1412.

Hückel Z 1931 Quanten theoretische Beiträge zum Benzolproblem. I Die Electron enkonfiguration des Benzols *Zeitschrift für Physik* **70**:203–286

Huzinga S 1965. Gaussian-type Functions for Polyatomic Systems I. *Journal of Chemical Physics* **42**:1293–1302.

Löwdin P-Q 1970. On the Orthogonality Problem. *Advances in Quantum Chemistry* **5**:185–199.

Mayer I 1983. Charge, Bond Order and Valence in the *Ab initio* SCF Theory. *Chemical Physics Letters* **97**:270–274

Mulliken R S 1955. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I *Journal of Chemical Physics* **23**.1833–1846.

Politzer P and J S Murray 1991. Molecular Electrostatic Potentials and Chemical Reactivity. In Lipkowitz K B and D B Boyd (Editors). *Reviews in Computational Chemistry* Volume 2. New York, VCH Publishers, pp 273–312.

Pople J A, D L Beveridge and P A Dobosh 1967 Approximate Self-Consistent Molecular Orbital Theory. V. Intermediate Neglect of Differential Overlap. *Journal of Chemical Physics* **47**:2026–2033.

Pople J A, D P Santry and G A Segal 1965. Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures. *Journal of Chemical Physics* **43**:S129–S135

Pople J A and G A Segal 1965. Approximate Self-Consistent Molecular Orbital Theory. II Calculations with Complete Neglect of Differential Overlap *The Journal of Chemical Physics* **43**:S136–S149

Pople J A and G A Segal 1966. Approximate Self-Consistent Molecular Orbital Theory. III. CNDO Results for AB_2 and AB_3 systems. *Journal of Chemical Physics* **44**:3289–3296

Reed A E, R B Weinstock and F Weinhold 1985. Natural Population Analysis. *Journal of Chemical Physics* **83**: 735–746.

Roothaan C C J 1951. New Developments in Molecular Orbital Theory *Reviews of Modern Physics* **23**:69–89

Slater J C 1930. Atomic Shielding Constants. *Physical Review* **36**.57–64

Smith G D and R L Jaffe 1996. Quantum Chemistry Study of Conformational Energies and Rotational Energy Barriers in *n*-Alkanes. *Journal of Physical Chemistry* **100**.18718–18724

Stewart J J P 1989a. Optimisation of Parameters for Semi-empirical Methods I. Method. *Journal of Computational Chemistry* **10**:209–220

Stewart J J P 1989b. Optimisation of Parameters for Semi-empirical Methods II. Applications. *Journal of Computational Chemistry* **10**.221–264

Thiel W and A A Voityuk 1994. Extension of MNDO to d Orbitals: Parameters and Results for Silicon. *Journal of Molecular Structure (Theochem)* **313**:141–154.

Wiberg K B and M A Murcko 1988. Rotational Barriers. 2. Energies of Alkane Rotamers. An Examination of Gauche Interactions *Journal of the American Chemical Society* **110**:8029–8038.

Wiberg K B and P R Rablen 1993. Comparison of Atomic Charges Derived via Different Procedures. *Journal of Computational Chemistry* **14**:1504–1518.

CHAPTER THREE

Advanced *ab initio* Methods, Density Functional Theory and Solid-state Quantum Mechanics

3.1 Introduction

In Chapter 2 we worked through the two most commonly used quantum mechanical models for performing calculations on ground-state 'organic'-like molecules, the *ab initio* and semi-empirical approaches. We also considered some of the properties that can be calculated using these techniques. In this chapter we will consider various advanced features of the *ab initio* approach and also examine the use of density functional methods. Finally, we will examine the important topic of how quantum mechanics can be used to study the solid state.

3.2 Open-shell Systems

The Roothaan–Hall equations are not applicable to open-shell systems, which contain one or more unpaired electrons. Radicals are, by definition, open-shell systems as are some ground-state molecules such as NO and O_2. Two approaches have been devised to treat open-shell systems. The first of these is *spin-restricted* Hartree–Fock (RHF) theory, which uses combinations of singly and doubly occupied molecular orbitals. The closed-shell approach that we have developed thus far is a special case of RHF theory. The doubly occupied orbitals use the same spatial functions for electrons of both α and β spin. The orbital expansion Equation (2.144) is employed together with the variational method to derive the optimal values of the coefficients. The alternative approach is the *spin-unrestricted Hartree–Fock* (UHF) theory of Pople and Nesbet [Pople and Nesbet 1954], which uses two distinct sets of molecular orbitals: one for electrons of α spin and the other for electrons of β spin. Two Fock matrices are involved, one for each type of spin, with elements as follows:

$$F_{\mu\nu}^{\alpha} = H_{\mu\nu}^{\text{core}} + \sum_{\lambda=1}^{K}\sum_{\sigma=1}^{K}[[P_{\lambda\sigma}^{\alpha} + P_{\lambda\sigma}^{\beta}](\mu\nu|\lambda\sigma) - P_{\lambda\alpha}^{\alpha}(\mu\lambda|\nu\sigma)] \tag{3.1}$$

$$F_{\mu\nu}^{\beta} = H_{\mu\nu}^{\text{core}} + \sum_{\lambda=1}^{K}\sum_{\sigma=1}^{K}[[P_{\lambda\sigma}^{\alpha} + P_{\lambda\sigma}^{\beta}](\mu\nu|\lambda\sigma) - P_{\lambda\alpha}^{\beta}(\mu\lambda|\nu\sigma)] \tag{3.2}$$

UHF theory also uses two density matrices, the full density matrix being the sum of these two:

$$P_{\mu\nu}^{\alpha} = \sum_{i=1}^{\alpha_{\text{occ}}} c_{\mu i}^{\alpha} c_{\nu i}^{\alpha} \qquad P_{\mu\nu}^{\beta} = \sum_{i=1}^{\beta_{\text{occ}}} c_{\mu i}^{\beta} c_{\nu i}^{\beta} \tag{3.3}$$

$$P_{\mu\nu} = P_{\mu\nu}^{\alpha} + P_{\mu\nu}^{\beta} \tag{3.4}$$

The summations in Equations (3.3) and (3 4) are over the occupied orbitals with α and β spin as appropriate. Thus, $\alpha_{\text{occ}} + \beta_{\text{occ}}$ equals the total number of electrons in the system. In a closed-shell Hartree–Fock wavefunction the distribution of electron spin is zero everywhere because the electrons are paired. In an open-shell system, however, there is an excess of electron spin, which can be expressed as the spin density, analogous to the electron density The spin density $\rho^{\text{spin}}(\mathbf{r})$ at a point $\mathbf{r}$ is given by:

$$\rho^{\text{spin}}(\mathbf{r}) = \rho^{\alpha}(\mathbf{r}) - \rho^{\beta}(\mathbf{r}) = \sum_{\mu=1}^{K}\sum_{\nu=1}^{K}[P_{\mu\nu}^{\alpha} - P_{\mu\nu}^{\beta}]\phi_{\mu}(\mathbf{r})\phi_{\nu}(\mathbf{r}) \tag{3 5}$$

Clearly, the UHF approach is more general and indeed the restricted Hartree–Fock approach is a special case of unrestricted Hartree–Fock. Figure 3.1 illustrates the conceptual difference between the RHF and the UHF models. Unrestricted wavefunctions are also the most appropriate way to deal with other problems such as molecules near the dissociation limit. The simplest example of this type of behaviour is the H_2 molecule, the ground state of which is a singlet with a bond length of approximately 0.75 Å. The restricted wavefunction is the appropriate Hartree–Fock wavefunction, with two paired electrons in a single spatial orbital. As the bond length increases towards the dissociation limit, this description is clearly inappropriate, for hydrogen is experimentally observed to dissociate to two

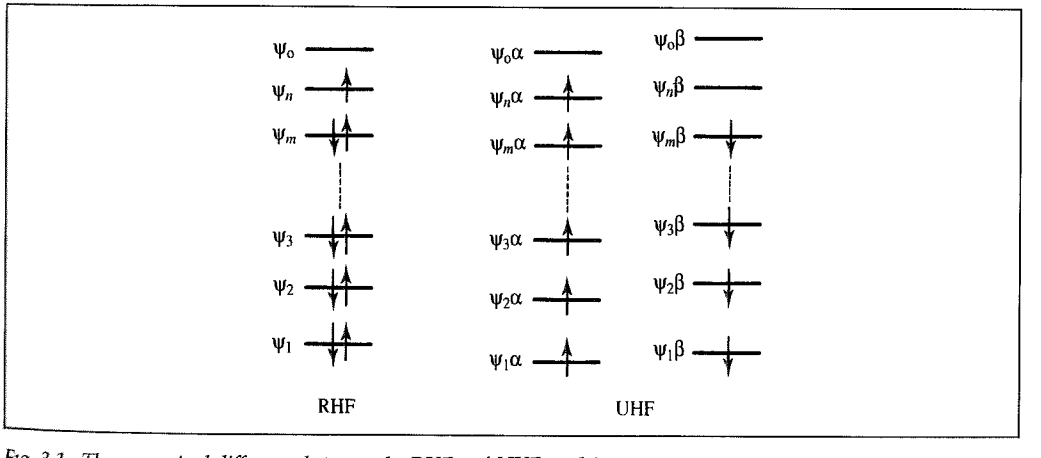

Fig. 3.1. The conceptual difference between the RHF and UHF models

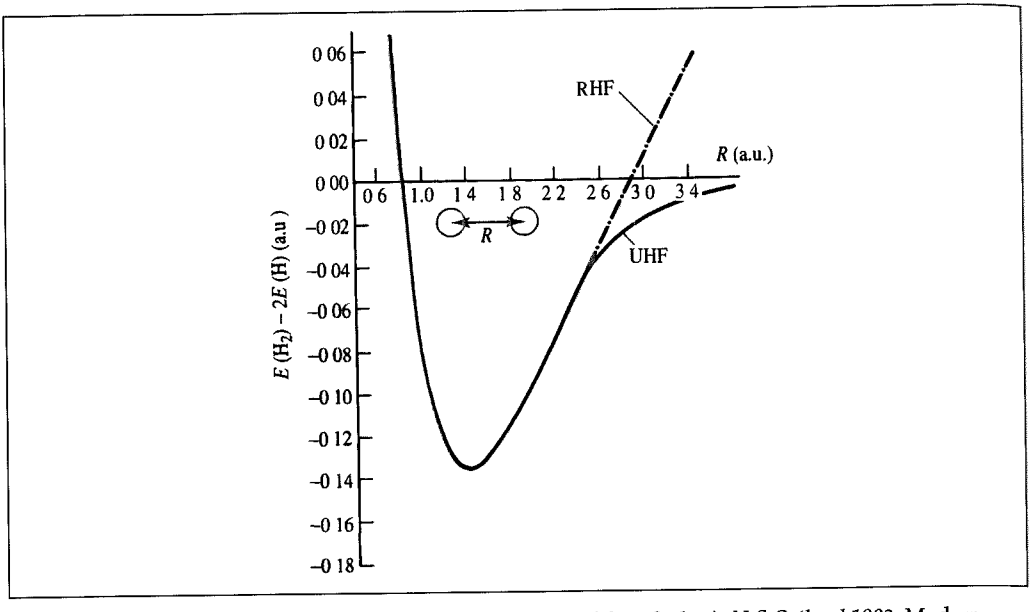

Fig. 3.2. UHF and RHF dissociation curves for H_2. (Figure adapted from Szabo A, N S Ostlund 1982. Modern Quantum Chemistry Introduction to Advanced Electronic Structure Theory. *New York, McGraw-Hill.)*

hydrogen atoms. This behaviour cannot be achieved using a restricted Hartree–Fock wavefunction, which requires the two electrons to occupy the same spatial orbital and leads to H^+ and H^-, but it is appropriately described by a UHF wavefunction. Beyond about 1.2 Å the 'correct' wavefunction for hydrogen must thus be obtained using UHF theory. The results obtained by calculating the potential energy curves of the hydrogen molecule using the RHF and UHF theories are shown in Figure 3.2. As can be seen, RHF theory gives a dissociation energy that is much too large, whereas the UHF theory shows the correct dissociation behaviour.

3.3 Electron Correlation

The most significant drawback of Hartree–Fock theory is that it fails to adequately represent electron correlation. In the self-consistent field method the electrons are assumed to be moving in an average potential of the other electrons, and so the instantaneous position of an electron is not influenced by the presence of a neighbouring electron. In fact, the motions of electrons are correlated and they tend to 'avoid' each other more than Hartree–Fock theory would suggest, giving rise to a lower energy. The correlation energy is defined as the difference between the Hartree–Fock energy and the exact energy. Neglecting electron correlation can lead to some clearly anomalous results, especially as the dissociation limit is approached. For example, an uncorrelated calculation would predict that the electrons in H_2 spend equal time on both nuclei, even when they are

infinitely separated. Hartree-Fock geometries and relative energies for equilibrium structures are often in good agreement with experiment and as many molecular modelling applications are concerned with species at equilibrium it might be considered that correlation effects are not so important. Nevertheless, there is increasing evidence that the inclusion of correlation effects is warranted, especially when quantitative information is required. Moreover, electron correlation is crucial in the study of dispersive effects (which we shall consider in Section 4.10.1), which play a major role in intermolecular interactions. Electron correlation is most frequently discussed in the context of *ab initio* calculations, but it should be noted that the effects of electron correlation are implicitly included in the semi-empirical methods because of the way in which they are parametrised. However, specific electron correlation methods have also been developed for use with the various levels of semi-empirical calculation; this in turn necessitates the modification of some parameters.

3.3.1 Configuration Interaction

There are a number of ways in which correlation effects can be incorporated into an *ab initio* molecular orbital calculation. A popular approach is configuration interaction (CI), in which excited states are included in the description of an electronic state. To illustrate the principle, let us consider a lithium atom. The ground state of lithium can be written $1s^2 2s^1$ (although we have used the conventional nomenclature here, we should remember that the wavefunction is really a Slater determinant). Excitation of the outer valence electron gives states such as $1s^2 3s^1$. A better description of the overall wavefunction is a linear combination of the ground and excited-state wavefunctions. If a Hartree–Fock calculation is performed with K basis functions then $2K$ spin orbitals are obtained. If these $2K$ spin orbitals are filled with N electrons ($N < 2K$) there will be $2K - N$ unoccupied, virtual orbitals. The wavefunction obtained from the single-determinant approach that we have considered thus far is expressed only in terms of the occupied orbitals. For example, a very simple calculation on H_2, using as a basis set just the 1s orbitals on each hydrogen, results in two molecular orbitals ($1\sigma_g$ and $1\sigma_u$). In the ground state, the $1\sigma_g$ orbital is filled with two electrons. An excited state can be generated by replacing one or more of the occupied spin orbitals with a virtual spin orbital. Possible excited states for the hydrogen molecule might thus include $1\sigma_g^1 \sigma_u^1$ and $1\sigma_u^2$ (in fact, the first of these two configurations cannot be combined with the ground state, as we shall see). In addition to the replacement of single spin orbitals by single virtual orbitals, two spin orbitals can be replaced by two virtual orbitals, three spin orbitals by three virtual orbitals, and so on. In general, the CI wavefunction can be written as:

$$\Psi = c_0 \Psi_0 + c_1 \Psi_1 + c_2 \Psi_2 + \cdots \tag{3.6}$$

Ψ_0 is the single-determinant wavefunction obtained by solving the Hartree–Fock equations. Ψ_1, Ψ_2, etc. are wavefunctions (expressed as determinants) that represent configurations derived by replacing one or more of the occupied spin orbitals by a virtual spin orbital. The energy of the system is then minimised in order to determine the coefficients c_0, c_1, etc., using a linear variational approach, just as for a single-determinant calculation. A CI

calculation thus involves an additional level of complexity; each configuration is written in terms of molecular orbitals, which in turn are expressed as a linear combination of basis functions. The number of integrals can become extremely large. The total number of ways to permute N electrons and K orbitals is $(2K!)/[N!(2K-N)!]$. This is a very large number for all except small values of K and N, which explains why it is not usual to consider all possibilities (termed *full configuration interaction*) except for very small systems. However, full CI is important because it is the most complete treatment possible within the limitations imposed by the basis set. In the limit of a complete basis set full CI becomes complete CI and virtually exact – but is generally considered impractical as at large K the number of Slater determinants increases exponentially with N as $K^N/N!$. It is common practice to limit the excited states considered. For example, in configuration interaction singles (CIS) only wavefunctions that differ from the Hartree–Fock wavefunction by a single spin orbital are included. The next levels of the theory involve double substitutions (configuration interaction doubles, CID) or both singles and double substitutions (configuration interaction singles and doubles, CISD). Even at the CIS or CID levels, the number of excited states to be included can be very large, and it may be desirable (or necessary) to restrict the spin orbitals that are involved in the substitutions. For example, only excitations involving the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) may be permitted. Alternatively, the orbitals corresponding to the inner electron core may be neglected (the 'frozen core' approximation). Some of these options are illustrated in Figure 3.3.

Not all excitations necessarily help to lower the energy; some determinants do not mix with the ground state. A consequence of *Brillouin's theorem* is that single excitations do not mix

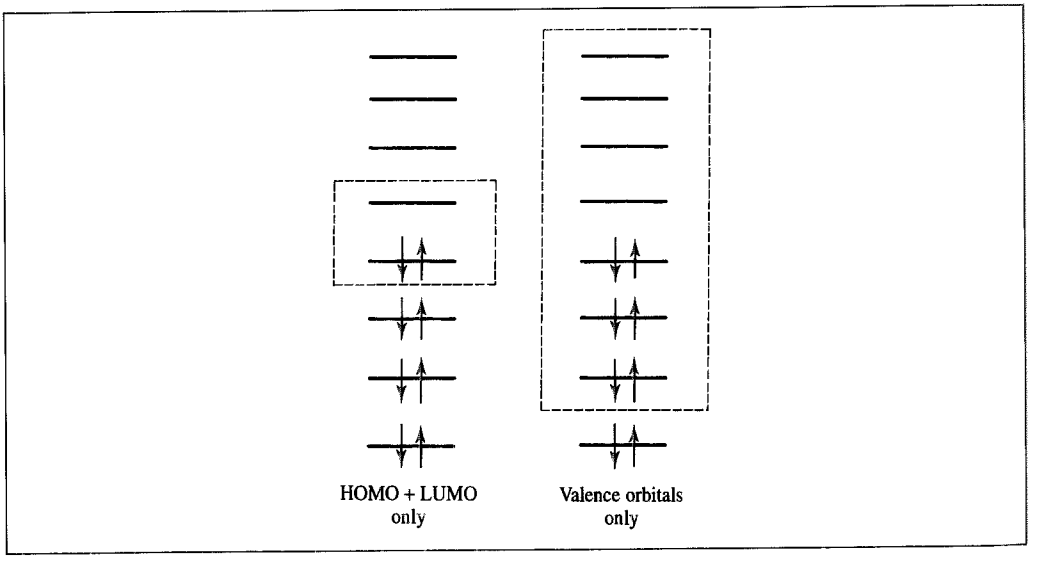

Fig. 3.3 Some of the ways in which excited-state wavefunctions can be included in a configuration interaction calculation (Figure adapted from Hehre W J, L Radom, P v R Schleyer and J A Hehre 1986 Ab initio Molecular Orbital Theory *New York, Wiley)*

directly with the single-determinant, ground-state wavefunction Ψ_0. It would therefore be anticipated that double excitations would be most important and that single excitations would have no effect on the energy of the ground state. However, the single excitations can interact with the double excitations, which in turn interact with Ψ_0, and so single excitations do have a small indirect effect on the energy. The determinants of triple and higher excitations also do not interact directly with Ψ_0 (though they may do indirectly via other levels of excitation). This is because the Hamiltonian contains elements involving at most interactions between pairs of electrons, and so if the Slater determinants differ by more than two electron functions, their integral over all space will be zero.

In a 'traditional' CI calculation the determinants in the expansion, Equation (3.6), are those obtained from a Hartree–Fock calculation; only the coefficients c_0, c_1, etc. are permitted to vary. Clearly, a better (i.e. lower-energy) wavefunction should be obtained if the coefficients of the basis functions themselves can vary as well as the coefficients of the determinants. This approach is known as the multiconfiguration self-consistent field method (MCSCF). MCSCF theory is considerably more complicated than the Roothaan–Hall equations and well beyond the scope of our discussion. One MCSCF technique that has attracted considerable attention is the complete active-space SCF method (CASSCF) of Roos [Roos *et al.* 1980]. CASSCF enables very large numbers of configurations to be included in the calculation by dividing the molecular orbitals into three sets: those which are doubly occupied in all configurations, those which are unoccupied in all configurations, and then all the remaining 'active' orbitals. The list of configurations is generated by considering all possible arrangements of the active electrons among the active orbitals.

A CI calculation is variational: the energy obtained is guaranteed to be greater than the 'true' energy. A drawback of CI calculations other than those performed at the full CI level is that they are not size consistent. Simply put, this means that the energy of a number N of non-interacting atoms or molecules is not equal to N times the energy of a single atom or molecule. Another consequence of size consistency is that, as the bond length in a diatomic molecule increases to infinity, so the energy of the system should become equal to the sum of the energies of the respective atoms. To illustrate why this lack of size consistency arises, consider CID calculations on Be_2 and on two beryllium atoms. The electronic configuration of Be is $1s^2 2s^2$ and so if we label the two atoms A and B, then the wavefunction for each of the two separated atoms will include the configuration $1s_A^2 2p_A^2 1s_B^2 2p_B^2$ ($\equiv 1s_A^2 1s_B^2 2p_A^2 2p_B^2$), in which two electrons have been promoted in each beryllium atom from the 2s to the 2p orbitals. This configuration represents a *quadruple* excitation for the beryllium dimer, which has the electronic configuration $1s_A^2 1s_B^2 2s_A^2 2s_B^2$. This quadruply excited configuration is not included in the CID wavefunction for the dimer, which is restricted to double excitations. In fact, the energy of a CI calculation including only doubly excited states is expected to scale in proportion to $\sqrt{N}$, where N is the number of non-interacting species present, rather than N. The Quadratic Configuration Interaction method (QCISD) was introduced to try to deal with this; it can be considered a size-consistent CISD theory [Pople *et al.* 1987]. The procedure involves the addition of higher excitation terms which are quadratic in their expansion coefficients. Higher still in theory is QCISD(T), in which an estimated contribution from the triple excitations can be incorporated, though with extra computational expense.

3.3.2 Many-body Perturbation Theory

Møller and Plesset proposed an alternative way to tackle the problem of electron correlation [Møller and Plesset 1934] Their method is based upon Rayleigh-Schrödinger perturbation theory, in which the 'true' Hamiltonian operator $\mathscr{H}$ is expressed as the sum of a 'zeroth-order' Hamiltonian $\mathscr{H}_0$ (for which a set of molecular orbitals can be obtained) and a perturbation, $\mathscr{V}$:

$$\mathscr{H} = \mathscr{H}_0 + \mathscr{V} \tag{3.7}$$

The eigenfunctions of the true Hamiltonian operator are Ψ_i with corresponding energies E_i. The eigenfunctions of the zeroth-order Hamiltonian are written $\Psi_i^{(0)}$ with energies $E_i^{(0)}$. The ground-state wavefunction is thus $\Psi_0^{(0)}$ with energy $E_0^{(0)}$. To devise a scheme by which it is possible to gradually improve the eigenfunctions and eigenvalues of $\mathscr{H}_0$ we can write the true Hamiltonian as follows:

$$\mathscr{H} = \mathscr{H}_0 + \lambda\mathscr{V} \tag{3.8}$$

λ is a parameter that can vary between 0 and 1; when λ is zero then $\mathscr{H}$ is equal to the zeroth-order Hamiltonian, but when λ is 1 then $\mathscr{H}$ equals its true value. The eigenfunctions Ψ_i and eigenvalues E_i of $\mathscr{H}$ are then expressed in powers of λ:

$$\Psi_i = \Psi_i^{(0)} + \lambda\Psi_i^{(1)} + \lambda^2\Psi_i^{(2)} + \cdots = \sum_{n=0} \lambda^n \Psi_i^{(n)} \tag{3.9}$$

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \cdots = \sum_{n=0} \lambda^n E_i^{(n)} \tag{3.10}$$

$E_i^{(1)}$ is the first-order correction to the energy, $E_i^{(2)}$ is the second-order correction, and so on. These energies can be calculated from the eigenfunctions as follows:

$$E_i^{(0)} = \int \Psi_i^{(0)} \mathscr{H}_0 \Psi_i^{(0)}\, d\tau \tag{3.11}$$

$$E_i^{(1)} = \int \Psi_i^{(0)} \mathscr{V} \Psi_i^{(0)}\, d\tau \tag{3.12}$$

$$E_i^{(2)} = \int \Psi_i^{(0)} \mathscr{V} \Psi_i^{(1)}\, d\tau \tag{3.13}$$

$$E_i^{(3)} = \int \Psi_i^{(0)} \mathscr{V} \Psi_i^{(2)}\, d\tau \tag{3.14}$$

To determine the corrections to the energy it is therefore necessary to determine the wavefunctions to a given order. In Møller–Plesset perturbation theory the unperturbed Hamiltonian $\mathscr{H}_0$ is the sum of the one-electron Fock operators for the N electrons:

$$\mathscr{H}_0 = \sum_{i=1}^{N} \mathscr{f}_i = \sum_{i=1}^{N} \left(\mathscr{H}^{\text{core}} + \sum_{j=1}^{N} (\mathscr{J}_i + \mathscr{K}_i) \right) \tag{3.15}$$

The Hartree–Fock wavefunction, $\Psi_0^{(0)}$, is an eigenfunction of $\mathscr{H}_0$, and the corresponding zeroth-order energy $E_0^{(0)}$ is equal to the sum or orbital energies for the occupied molecular

orbitals:

$$E_0^{(0)} = \sum_{i=1}^{\text{occupied}} \varepsilon_i \tag{3.16}$$

In order to calculate higher-order wavefunctions we need to establish the form of the perturbation, $\mathscr{V}$. This is the difference between the 'real' Hamiltonian $\mathscr{H}$ and the zeroth-order Hamiltonian, $\mathscr{H}_0$. Remember that the Slater determinant description, based on an orbital picture of the molecule, is only an approximation. The true Hamiltonian is equal to the sum of the nuclear attraction terms and electron repulsion terms:

$$\mathscr{H}_0 = \sum_{i=1}^{N} (\mathscr{H}^{\text{core}}) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{r_{ij}} \tag{3.17}$$

Hence the perturbation $\mathscr{V}$ is given by:

$$\mathscr{V} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{r_{ij}} - \sum_{j=1}^{N} (\mathscr{J}_j + \mathscr{K}_j) \tag{3.18}$$

The first-order energy $E_0^{(1)}$ is given by:

$$E_0^{(1)} = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{r_{ij}} [(ii|jj) - (ij|ij)] \tag{3.19}$$

The sum of the zeroth-order and first-order energies thus corresponds to the Hartree–Fock energy (compare with Equation (2.110), which gives the equivalent result for a closed-shell system):

$$E_0^{(0)} + E_0^{(1)} = \sum_{i=1}^{N} \varepsilon_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} [(ii|jj) - (ij|ij)] \tag{3.20}$$

To obtain an improvement on the Hartree–Fock energy it is therefore necessary to use Møller–Plesset perturbation theory to at least second order. This level of theory is referred to as MP2 and involves the integral $\int \Psi_0^{(0)} \mathscr{V} \Psi_0^{(1)} \, d\tau$. The higher-order wavefunction $\Psi_0^{(1)}$ is expressed as linear combinations of solutions to the zeroth-order Hamiltonian:

$$\Psi_0^{(1)} = \sum_j c_j^{(1)} \Psi_j^{(0)} \tag{3.21}$$

The $\Psi_j^{(0)}$ in Equation (3.21) will include single, double, etc. excitations obtained by promoting electrons into the virtual orbitals obtained from a Hartree–Fock calculation. The second-order energy is given by:

$$E_0^{(2)} = \sum_{i}^{\text{occupied}} \sum_{j>1} \sum_{a}^{\text{virtual}} \sum_{b>a} \frac{\iint d\tau_1 \, d\tau_2 \, \chi_i(1)\chi_j(2) \left(\frac{1}{r_{12}}\right) [\chi_a(1)\chi_b(2) - \chi_b(1)\chi_a(2)]}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j} \tag{3.22}$$

These integrals will be non-zero only for double excitations, according to the Brillouin theorem. Third- and fourth-order Møller–Plesset calculations (MP3 and MP4) are also

available as standard options in many *ab initio* packages. For the fourth-order calculations single, triple and quadruple excitations will also contribute. As the triple substitutions are most difficult to perform computationally a partial theory that involves just single, double and quadruple substitutions (MP4SDQ) is a popular alternative.

The advantage of many-body perturbation theory is that it is size-independent, unlike configuration interaction - even when a truncated expansion is used. However, Møller-Plesset perturbation theory is not variational and can sometimes give energies that are lower than the 'true' energy. Møller-Plesset calculations are computationally intensive and so their use is often restricted to 'single-point' calculations at a geometry obtained using a lower level of theory. They are at present the most popular way to incorporate electron correlation into molecular quantum mechanical calculations, especially at the MP2 level. A Møller-Plesset calculation is specified using the level of theory used (e.g. MP2, MP3) together with the basis set. Thus MP2/6-31G* indicates a second-order Møller-Plesset calculation with the 6-31G* basis set.

Certain properties benefit more from the use of correlation methods than others do. For example, a single-determinant Hartree-Fock method and a reasonable basis set give geometrical parameters often very close (bond lengths within 0.01–0.02 Å and angles within 1–2°) to the experimental values. This contrasts with the situation for processes which result in the unpairing of electrons. A simple example is the bond dissociation energy of H_2, for which the Hartree-Fock limit is 84 kcal/mol. MP2, MP3 and MP4 calculations using the 6-31G** basis set give results of 101, 105 and 106 kcal/mol, respectively, for this process, much closer to the experimental value of 109 kcal/mol. In these and similar situations, electron correlation is often advised, if the computational resources permit. However, one class of reactions can be well described using single-determinant Hartree-Fock theory. These are known as *isodesmic reactions*, which are transformations in which the number of electron pairs is constant and the chemical bond types are conserved. Such reactions would be expected to benefit from a judicious cancellation of errors as only the environment of the bonds has changed. Examples of isodesmic reactions are:

$$CH_3CH_2CH_3 + CH_4 \rightarrow 2CH_3CH_3$$

$$CF_3CHO + CH_4 \rightarrow CF_3H + CH_3CHO$$

$$CH_3CH{=}C{=}O + 2CH_4 \rightarrow CH_3CH_3 + CH_2{=}CH_2 + H_2C{=}O$$

Even at the STO-3G level quite respectable results can often be obtained.

In an attempt to deal with some of the shortcomings of even the correlated methods a number of correction factors have been developed. The Gaussian-*n* procedures [Pople *et al.* 1989, Curtiss *et al.* 1991, 1998] represent an attempt to develop a protocol for the accurate calculation of various properties such as atomisation energies, ionisation potentials, electron affinities and proton affinities for atoms and molecules containing first-row and second-row elements. Currently, the most recent member of this series is Gaussian-3 (G3) theory [Curtiss *et al.* 1998]. The G3 method involves a defined sequence of calculations involving geometry optimisation first at the Hartree-Fock level with the 6-31G* basis set and then at the MP2/6-31G* level. A single-point calculation is next carried out using this geometry with the full MP4 method (singles, doubles, triples and quadruples). This energy is then

refined through a series of corrections, which deal with the need for higher polarisation functions, for correlation effects beyond fourth-order perturbation theory (i.e. QCISD(T)) and for larger basis set effects. These correction factors are combined, together with a zero-point energy derived from a series of scaled harmonic frequencies determined from the first, HF/6-31G', geometry optimisation, to give the final G3 energy. When tested on 299 experimental energies the overall average absolute deviation from experiment was 1.02 kcal/mol, with the average deviations for the four different types of data being 0.94 kcal/mol for the enthalpies of formation (148 values), 1.13 kcal/mol for the ionisation energies (85 values), 1.00 kcal/mol for electron affinities (58 values) and 1.34 kcal/mol for proton affinities (eight values). Detailed examination of the results can help to identify systems requiring most attention in subsequent developments of the theory. For example, the enthalpy of formation of both SO_2 and PF_3 have large negative deviations from experiment, perhaps due to the need for a larger basis set to describe the bonding in these molecules. Likewise some of the strained hydrocarbon ring systems (cyclopropene, cyclobutene and bicyclobutane) also show relatively large deviations.

The G3 method is still rather computationally intensive and so some efforts have been made to reduce the computational requirements whilst retaining an acceptable level of error. The G3(MP2) variant [Curtiss *et al.* 1999] replaces the MP4 calculations (which are particularly time-consuming), with comparable calculations at the MP2 level. This leaves the QCISD(T) stage as the most demanding step. The average absolute deviation of the energies calculated using the G3(MP2) method was 1.89 kcal/kmol on the entire 299 test systems, a significantly less accurate result than that of the full G3 method, but still noteworthy.

3.4 Practical Considerations When Performing *ab initio* Calculations

Ab initio calculations can be extremely time-consuming, especially when using the higher levels of theory or when the nuclei are free to move, as in a minimisation calculation (see Chapter 5). Various 'tricks' have been developed which can significantly reduce the computational effort involved. Many of these options are routinely available in the major software packages and are invoked by the specification of simple keywords. One common tactic is to combine different levels of theory for the various stages of a calculation. For example, a lower level of theory can be used to provide the initial guess for the density matrix prior to the first SCF iteration. Lower levels of theory can also be used in other ways. Suppose we wish to determine some of the electronic properties of a molecule in a minimum energy structure. Energy minimisation requires that the nuclei move and is typically performed in a series of steps, at each of which the energy (and frequently the gradient of the energy) must be calculated. Minimisation is therefore a computationally expensive procedure, particularly when performed at the high level of theory. To reduce this computational burden a lower level of theory can be employed for the geometry optimisation. A 'single-point' calculation using a high level of theory is then performed at the geometry so obtained to give a wavefunction from which the properties are determined. The assumption here of course is that the geometry does not change much between the two levels of

theory. Such calculations are denoted by slashes (/). For example, a calculation that is described as '6-31G*/STO-3G' indicates that the geometry was determined using the STO-3G basis set and the wavefunction was obtained using the 6-31G* basis set. Two slashes are used when each calculation is itself described using a slash, such as when electron correlation methods are used. For example, 'MP2/6-31G*//HF/6-31G*' indicates a geometry optimisation using a Hartree–Fock calculation with a 6-31G* basis set followed by a single-point calculation using the MP2 method for incorporating electron correlation, again using a 6-31G* basis set.

3.4.1 Convergence of Self-consistent Field Calculations

In an SCF calculation the wavefunction is gradually refined until self-consistency is achieved For closed-shell ground-state molecules this is usually quite straightforward and the energy converges after a few cycles. However, in some cases convergence is a problem, and the energy may oscillate from one iteration to the next or even diverge rapidly. Various methods have been proposed to deal with such situations. A simple strategy is to use an average set of orbital coefficients rather than the set obtained from the immediately preceding iteration. The coefficients in this average set can be weighted according to the energies of each iteration. This tends to weed out those coefficients that give rise to higher energies.

The initial guess of the density matrix may influence the convergence of the SCF calculation; a null matrix is the simplest approach, but better results may be obtained by using a density matrix from a calculation performed at a lower level of theory. For example, the density matrix from a semi-empirical calculation may be used as the starting point for an *ab initio* calculation. Conversely, such an approach may itself lead to problems if there is a significant difference between the density matrices for the lower and higher levels of theory.

A more sophisticated method that is often very successful is Pulay's direct inversion of the iterative subspace (DIIS) [Pulay 1980]. Here, the energy is assumed to vary as a quadratic function of the basis set coefficients. In DIIS the coefficients for the next iteration are calculated from their values in the previous steps. In essence, one is predicting where the minimum in the energy will lie from a knowledge of the points that have been visited and by assuming that the energy surface adopts a parabolic shape.

3.4.2 The Direct SCF Method

An *ab initio* calculation can be logically considered to involve two separate stages. First, the various one- and two-electron integrals are calculated. This is a computationally intensive task and considerable effort has been expended finding ways to make the calculation of the integrals as efficient as possible. In the second stage, the wavefunction is determined using the variation theorem. In a 'traditional' SCF calculation all of the integrals are first calculated and stored on disk, to be retrieved later during the SCF calculation as required. The number of integrals to be stored may run into millions and this inevitably leads to delays in accessing the data, particularly as the retrieval of information from a disk requires

physical movement of the read head and so is slow. Modern computers (both workstations and supercomputers) have much faster (and cheaper) processing units, and many of these machines also have a substantial amount of internal memory, which can be accessed in a fraction of the time it takes to read data from the disk. In a direct SCF calculation, the integrals are not stored on the disk but are kept in memory or recalculated when required [Almlöf *et al.* 1982].

A much-quoted 'fact' is that *ab initio* calculations scale as the fourth power of the number of basis functions for ground-state, closed-shell systems. This scaling factor arises because each two-electron integral $(\mu\nu|\lambda\sigma)$ involves four basis functions, so the number of two-electron integrals would be expected to increase in proportion to the fourth power of the number of basis functions. In fact, the number of such integrals is not exactly equal to the fourth power of the number of basis functions because many of the integrals are related by symmetry. We can calculate exactly the number of two-electron integrals that are required in a Hartree–Fock *ab initio* calculation as follows. There are seven different types of two-electron integral:

1. $(ab|cd) \equiv (ab|dc) \equiv (ba|cd) \equiv (ba|dc) \equiv (cd|ab) \equiv (cd|ba) \equiv (dc|ab) \equiv (dc|ba)$
2. $(aa|bc) \equiv (aa|cb) \equiv (bc|aa) \equiv (cb|aa)$
3. $(ab|ac) \equiv (ab|ca) \equiv (ba|ac) \equiv (ba|ca) \equiv (ac|ab) \equiv (ac|ba) \equiv (ca|ab) \equiv (ca|ba)$
4. $(aa|bb) \equiv (bb|aa)$
5. $(ab|ab) \equiv (ab|ba) \equiv (ba|ab) \equiv (ba|ba)$
6. $(aa|ab) \equiv (aa|ba) \equiv (ab|aa) \equiv (ba|aa)$
7. $(aa|aa)$

For a basis set with K basis functions, there are $K(K-1)(K-2)(K-3)$ integrals of type $(ab|cd)$, but due to symmetry only one-eighth of these are unique as shown. Similarly, there are $2K(K-1)(K-2)$ of type (2); $4K(K-1)(K-2)$ of type (3); $K(K-1)$ of type (4); $2K(K-1)$ of type (5); $4K(K-1)$ of type (6) and K of type 7. Thus, a basis set with 200 functions has a total of 202 015 050 unique two-electron integrals. For all but the smallest of basis sets most integrals are of type (1) which is why an *ab initio* problem is often considered to scale as $K^4/8$ ($200^4/8 = 200\,000\,000$). Including electron correlation adds significantly to the computational cost; for example, MP2 calculations scale as the fifth power of the number of basis functions. Electron correlation methods may also require significantly more memory and disk than the comparable SCF calculation; the higher levels scale as the sixth power, and in QCISD(T), one part of the calculation is seventh order.

In practice, *ab initio* calculations often scale as a significantly smaller power than four. It is found that in favourable cases the computational cost of a direct SCF calculation on a large molecule scales as approximately the *square* of the number of basis functions used. This significant reduction (from four to two) is due to several factors. We have already noted some of the ways in which a carefully chosen basis set can reduce the computational effort, for example by making many of the integrals (particularly the two-electron integrals) identical by using the same Gaussian exponents for s and p orbitals in the same shell. Another way in which the calculation time can be significantly reduced is to exploit any symmetry of the system. Many isolated molecules contain symmetry elements such as centres of inversion and mirror planes, information which can be used to reduce the

computational effort required. In the case of an *ab initio* calculation that scales as the fourth power of the number of basis functions then a four-fold reduction in the number of atoms can (in principle at least) result in the computational time being reduced by about 250 times. The most effective way to reduce the computational effort is to identify integrals which are so small that ignoring them (i.e. setting them to zero) will not affect the results. The number of 'important' integrals is believed to scale as $K^2 \ln K$. The negligible integrals are determined by calculating an upper limit for each integral. This can be done rapidly and so those integrals that are guaranteed to be negligible can be identified and so ignored. The cutoff value which determines whether an integral is explicitly calculated or is set to zero can vary from one program to another, so it is always useful to check its value if different programs give different results for a given calculation.

3.4.3 Calculating Derivatives of the Energy

Considerable effort has been spent devising efficient ways of directly calculating the first and second derivatives of the energy with respect to the nuclear coordinates. Derivatives are primarily used during minimisation procedures for finding equilibrium structures (the first derivative of the energy with respect to its coordinates equals the force on an atom) and are also used by methods which locate transition structures and determine reaction pathways.

A self-consistent field wavefunction (and thus its energy) can be considered a complicated function of the nuclear coordinates, basis functions and basis function coefficients (and, for a CI calculation, the coefficients of single determinantal wavefunctions). In order to determine the first, second, etc. derivatives of the energy with respect to the nuclear coordinates [Pulay 1977] it is necessary to consider not only how the energy depends directly on the nuclear coordinates but also whether there is an indirect dependence via other parameters. Indeed, it is only the one-electron part of the Hamiltonian that depends directly upon the nuclear coordinates ($H^{core}(1)$, Equation (2.125)), to which is added an internuclear Coulomb repulsion term. For the other parameters the derivative with respect to the nuclear coordinates is generally determined via the chain rule (for first derivatives). For example, for a generic nuclear coordinate q_i and a generic parameter x_j we can write:

$$\frac{\partial E}{\partial q_i} = \frac{\partial E}{\partial x_j} \frac{\partial x_j}{\partial q_i} \tag{3.23}$$

In Equation (3.23) q_i would be the x, y or z coordinate of an atom and x_j would be a parameter such as a basis function coefficient or a basis function exponent. An important result is that the terms involving variationally determined parameters (such as basis function coefficients) are equal to zero; the energy is a minimum when $(\partial E/\partial c_j)$ is zero. This greatly reduces the computational effort. Most of the numerical work in calculating the gradient is due to the various basis set parameters (e.g. orbital centres and exponents) which require the derivatives of the various electron integrals. For Gaussian basis sets these derivatives can be obtained analytically and indeed it is relatively straightforward to obtain first derivatives for many levels of theory. The time taken to calculate the derivatives is comparable to that required for the calculation of the total energy. Second (and

higher) derivatives are more difficult and expensive to calculate, even at the lower levels of theory.

A possible alternative approach to the calculation of forces is via the use of the Hellmann–Feynman theorem. If Ψ is an exact wavefunction of a Hamiltonian H with energy E then this theorem states that the derivative of E with respect to some parameter P can be written:

$$\frac{\partial E}{\partial P} = \left\langle \frac{\partial H}{\partial P} \right\rangle \tag{3.24}$$

In the case of the derivative with respect to some nuclear coordinate q_i, we would consider the exact force and the Hellmann–Feynmann force to be equal:

$$\frac{\partial}{\partial q_i}\langle \Psi | H | \Psi \rangle = \left\langle \Psi \left| \frac{\partial H}{\partial q_i} \right| \Psi \right\rangle \tag{3.25}$$

Unfortunately, this only holds for the exact wavefunction and certain other types of wavefunction (such as at the Hartree–Fock limit). Moreover, even though the Hellmann–Feynman forces are much easier to calculate they are very unreliable, even for accurate wavefunctions, giving rise to spurious forces (often referred to as 'Pulay forces' [Pulay 1987]).

3.4.4 Basis Set Superposition Error

Suppose we wish to calculate the energy of formation of a bimolecular complex, such as the energy of formation of a hydrogen-bonded water dimer. Such complexes are sometimes referred to as 'supermolecules'. One might expect that this energy value could be obtained by first calculating the energy of a single water molecule, then calculating the energy of the dimer, and finally subtracting the energy of the two isolated water molecules (the 'reactants') from that of the dimer (the 'products'). However, the energy difference obtained by such an approach will invariably be an overestimate of the true value. The discrepancy arises from a phenomenon known as *basis set superposition error* (BSSE). As the two water molecules approach each other, the energy of the system falls not only because of the favourable intermolecular interactions but also because the basis functions on each molecule provide a better description of the electronic structure around the other molecule. It is clear that the BSSE would be expected to be particularly significant when small, inadequate basis sets are used (e.g. the minimal basis STO-*n*G basis sets) which do not provide for an adequate representation of the electron distribution far from the nuclei, particularly in the region where non-covalent interactions are strongest. One way to estimate the basis set superposition error is via the counterpoise correction method of Boys and Bernardi, in which the entire basis set is included in all calculations [Boys and Bernardi 1970]. Thus, in the general case:

$$\mathrm{A} + \mathrm{B} \rightarrow \mathrm{AB} \tag{3.26}$$

$$\Delta E = E(\mathrm{AB}) - [E(\mathrm{A} + E(\mathrm{B})] \tag{3.27}$$

The calculation of the energy of the individual species A is performed in the presence of 'ghost' orbitals of B; that is, without the nuclei or electrons of B. A similar calculation is

performed for B using ghost orbitals on A. An alternative approach is to use a basis set in which the orbital exponents and contraction coefficients have been optimised for molecular calculations rather than for atoms. The relevance of the basis set superposition error and its dependence upon the basis set and the level of theory employed (i.e. SCF or with electron correlation) remains a subject of much research.

3.5 Energy Component Analysis

The interaction between atoms and molecules can vary from the weak attraction between a pair of closed-shell atoms (e.g. two rare gas atoms in a molecular beam) to the large energy associated with the formation of a chemical bond. Intermediate between these two extremes are interactions due to hydrogen bonding or electron donor–acceptor processes. In these intermediate cases it is often difficult to determine what factors are important in contributing to the interaction. For example, what can a hydrogen bond be ascribed to?

Morokuma analysis is a method for decomposing the energy change on formation of an intermolecular complex into five components: electrostatic, polarisation, exchange repulsion, charge transfer and mixing [Morokuma 1977]. Suppose we have performed *ab initio* SCF calculations on two molecules, X and Y, and on the intermolecular complex (or 'supermolecule') XY. The wavefunctions obtained can be written $A\Psi_X^0$, $A\Psi_Y^0$ and $A\Psi_{XY}^0$. 'A' indicates the use of an antisymmetrised wavefunction (e.g. a Slater determinant). The sum of the energies of the isolated molecules is E_0 and the energy of the supermolecule is E_4 (we follow the original notation of Morokuma). The interaction energy ΔE is thus given by $E_4 - E_0$. The five components are calculated as follows.

The electrostatic contribution equals the interaction between the unperturbed electron distributions of the two isolated species, A and B. It is identical to the classical Coulomb interaction and equals the difference $E_1 - E_0$, where E_1 is the energy associated with the product of the two individual wavefunctions, Ψ_1:

$$\Psi_1 = A\Psi_A^0 A\Psi_B^0 \tag{3.28}$$

The electronic distributions of both X and Y will be changed by the presence of the other molecule. These polarisation effects cause a dipole to be induced in (say) molecule Y due to the charge distribution in molecule X and vice versa. Polarisation also affects the higher-order multipoles. To calculate the polarisation contribution we first calculate molecular wavefunctions Ψ_A and Ψ_B in the presence of the other molecule. The energy of the wavefunction Ψ_2 is determined as E_2, where Ψ_2 is:

$$\Psi_2 = A\Psi_X A\Psi_Y \tag{3.29}$$

The polarisation contribution equals $E_2 - E_1$ and is always attractive.

In determining Ψ_1 and Ψ_2, no electron exchange interactions are considered. The overlap between the electron distributions of X and Y at short range causes a repulsion because to bring together electrons with the same spin into the same region of space ultimately leads to a violation of the Pauli principle.

The exchange repulsion is calculated as $E_3 - E_1$, where E_3 is the energy of the wavefunction Ψ_3:

$$\Psi_3 = A(\Psi_X^0 \cdot \Psi_Y^0) \tag{3.30}$$

Ψ_3 is derived from the undistorted wavefunctions of X and Y but the exchange of electrons is permitted. The exchange term is always repulsive.

The charge transfer term arises from the transfer of charge (i.e. electrons) from occupied molecular orbitals on one molecule to unoccupied orbitals on the other molecule. This contribution is calculated as the difference between the energy of the supermolecule XY when this charge transfer is specifically allowed to occur, and an analogous calculation in which it is not.

The Morokuma formalism also requires an additional, 'mixing' or 'coupling' term to be included This equals the difference between the total SCF difference, ΔE, and the sum of the four contributions (electrostatic, polarisation, exchange repulsion and charge transfer). The mixing term has little physical significance and is used because the four components do not completely account for the entire interaction energy (it is a fudge factor!). Fortunately, it is often relatively small.

Morokuma studied a number of hydrogen-bonded complexes using this scheme in order to assess the contribution from each component. The systems studied were typically of intermolecular complexes involving small molecules such as H_2O, HF and NH_3. In addition, Morokuma and his colleagues also examined a series of electron donor–acceptor complexes such as H_3N–BF_3, OC–BH_3, HF–CIF and benzene–$OC(CN)_2$. He also studied the basis-set dependence of the results and observed that the energy components were more sensitive than the energy differences. For example, a minimal STO-3G basis set overestimates the charge transfer contribution, whereas double zeta basis sets tend to exaggerate the electrostatic interaction.

3.5.1 Morokuma Analysis of the Water Dimer

The water dimer $(H_2O)_2$ has been subject to perhaps the closest scrutiny of all hydrogen-bonded complexes. A variety of stable geometries are available to the water dimer, in which one or more hydrogen bonds are present. There has been considerable debate over the relative energies of these structures and even some dispute over which structures are actually at minimum points on the energy surface [Smith *et al.* 1990]. As might be expected, the results depend upon the basis set used. A linear geometry is observed experimentally and is also predicted to be the most stable structure by *ab initio* calculations with a wide variety of basis sets (see Figure 3.4). Using a 6-31G** basis set, Umeyama and Morokuma calculated that the −5.6 kcal/mol stabilisation energy was composed of −7.5 kcal/mol electrostatic stabilisation, 4.3 kcal/mol exchange repulsion, −0.5 kcal/mol polarisation and −1.8 kcal/mol charge transfer [Umeyama and Morokuma 1977]. The 'mixing term' contributed −0.1 kcal/mol. Thus the hydrogen bond in the water dimer was considered to arise primarily from electrostatic effects with a smaller charge transfer contribution. Morokuma and Umeyama also extended their analysis of charge transfer to investigate whether this was due to transfer from the proton donor to the acceptor, or from acceptor

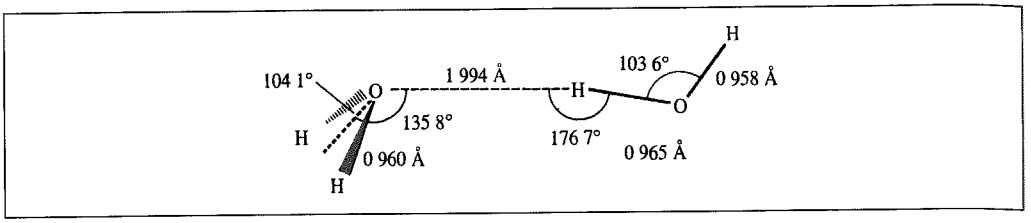

Fig. 3.4 The linear structure of the water dimer [Smith et al *1990].*

to donor. The results showed that approximately 90% of the charge transfer resulted from proton acceptor to proton donor transfer.

Morokuma analysis was widely used in the years after its introduction; it is less popular now as some problems have been encountered when trying to interpret the results with the larger basis sets that are feasible with today's faster computers and improved algorithms. In particular, when diffuse basis sets are used then there is a substantial amount of intermolecular overlap even at relatively large distances, which can make it difficult to factor out the different components. Nevertheless, the approach is certainly a useful way to assess the major causes of a particular type of intermolecular interaction, if only to provide a qualitative picture.

3.6 Valence Bond Theories

An entirely different way to treat the electronic structure of molecules is provided by valence bond theory, which was developed at about the same time as the molecular orbital approach. However, valence bond theory was not so amenable to calculations on large molecules, and molecular orbital theory came to dominate electronic structure theory for such systems. Nevertheless, valence bond theories are often considered to be more appropriate for certain types of problem than molecular orbital theory, especially when dealing with processes that involve bonds being broken and/or formed. Recall from Figure 3.2 that a self-consistent field wavefunction gives a wholly inaccurate picture for the dissociation of H_2; by contrast, the correct dissociation behaviour is naturally built into valence bond theories.

Valence bond theory is usually introduced using the famous Heitler–London model of the hydrogen molecule [Heitler and London 1927] This model considers two non-interacting hydrogen atoms (a and b) in their ground states that are separated by a long distance. The wavefunction for this system is:

$$\Psi = \phi_{1sa}(1)\phi_{1sb}(2) \tag{3.31}$$

As the two hydrogen atoms approach to form a hydrogen molecule, such a wavefunction is inappropriate as it implies that electron 1 remains confined to orbital 1sa and electron 2 to orbital 1sb. This clearly violates the indistinguishability principle, and so a linear combination is used

$$\Psi_{vb} \propto \phi_{1sa}(1)\phi_{1sb}(2) + \phi_{1sa}(2)\phi_{1sb}(1) \tag{3.32}$$

The corresponding molecular orbital function for this system is:

$$\Psi_{\mathrm{mo}} \propto \phi_{1\mathrm{sa}}(1)\phi_{1\mathrm{sb}}(2) + \phi_{1\mathrm{sa}}(2)\phi_{1\mathrm{sb}}(1) + \phi_{1\mathrm{sa}}(1)\phi_{1\mathrm{sa}}(2) + \phi_{1\mathrm{sb}}(1)\phi_{1\mathrm{sb}}(2) \qquad (3\,33)$$

The additional terms in the molecular orbital wavefunction correspond to states with the two electrons in the same orbital, which endows ionic character to the bond (H^+H^-). The valence bond wavefunction does not include any ionic character and in fact it correctly describes the dissociation into two hydrogen atoms. The simple valence bond and molecular orbital pictures in Equations (3.32) and (3.33) are extremes, with the 'true' wavefunction being somewhere in the middle. The valence bond representation can be improved by including a degree of ionic character as follows:

$$\Psi_{\mathrm{vb}} \propto \phi_{1\mathrm{sa}}(1)\phi_{1\mathrm{sb}}(2) + \phi_{1\mathrm{sa}}(2)\phi_{1\mathrm{sb}}(1) + \lambda[\phi_{1\mathrm{sa}}(1)\phi_{1\mathrm{sa}}(2) + \phi_{1\mathrm{sb}}(1)\phi_{1\mathrm{sb}}(2)] \qquad (3.34)$$

λ is a parameter that can be varied to give the 'correct' amount of ionic character. Another way to view the valence bond picture is that the incorporation of ionic character corrects the overemphasis that the valence bond treatment places on electron correlation. The molecular orbital wavefunction underestimates electron correlation and requires methods such as configuration interaction to correct for it. Although the presence of ionic structures in species such as H_2 appears counterintuitive to many chemists, such species are widely used to explain certain other phenomena such as the ortho/para or meta directing properties of substituted benzene compounds under electrophilic attack. Moverover, it has been shown that the ionic structures correspond to the deformation of the atomic orbitals when they are involved in chemical bonds.

One widely used valence bond theory is the generalised valence bond (GVB) method of Goddard and co-workers [Bobrowicz and Goddard 1977]. In the simple Heitler–London treatment of the hydrogen molecule the two orbitals are the non-orthogonal atomic orbitals on the two hydrogen atoms. In the GVB theory the analogous wavefunction is written:

$$\Psi_{\mathrm{GVB}} \propto u(1)\nu(2) + u(2)\nu(1) \qquad (3.35)$$

u and ν are non-orthogonal orbitals that are each expressed as a basis set expansion with the coefficients being variationally optimised to minimise the energy. The construction of the wavefunction from orbitals that are not necessarily orthogonal is characteristic of many valence bond theories and complicates the computational problem. The GVB approach is particularly successful for describing the electronic nature of systems as they approach dissociation.

Another approach is spin-coupled valence bond theory, which divides the electrons into two sets: 'core' electrons, which are described by doubly occupied orthogonal orbitals, and 'active' electrons, which occupy singly occupied non-orthogonal orbitals. Both types of orbital are expressed in the usual way as a linear combination of basis functions. The overall wavefunction is completed by two spin functions; one that describes the coupling of the spins of the core electrons and one that deals with the active electrons. The choice of spin function for these active electrons is a key component of the theory [Gerratt *et al.* 1997]. One of the distinctive features of this theory is that a considerable amount of chemically significant electronic correlation is incorporated into the wavefunction, giving an accuracy comparable to CASSCF. An additional benefit is that the orbitals tend to be

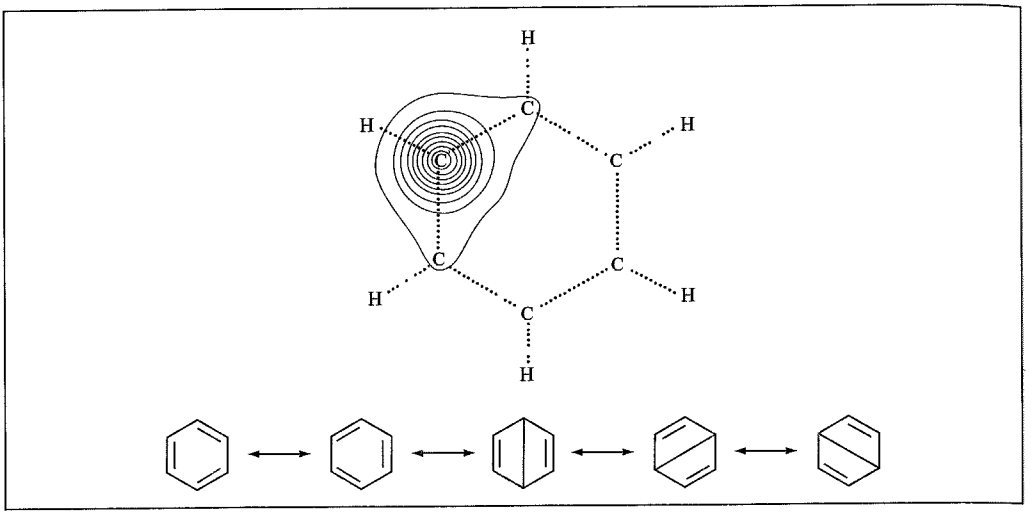

Fig 3 5· π orbital for benzene obtained from spin-coupled valence bond theory (Figure redrawn from Gerratt J, D L Cooper, P B Karadakov and M Raimondi 1997 Modern valence bond theory Chemical Society Reviews *87 100) The figure also shows the two Kekulé and three Dewar benzene forms which contribute to the overall wavefunction; each Kekulé form contributes approximately 40.5% and each Dewar form approximately 6.4%*

localised, closely resembling atomic or hybrid atomic orbitals, and consequently very visual. Various chemical phenomena have been examined using this approach, including dissociation reactions and hypervalence. One particularly interesting study was of the π system of benzene [Cooper *et al.* 1986]. This calculation resulted in six orbitals, each localised on one of the carbon atoms in the ring, though with some deformations towards neighbouring atoms (Figure 3.5). Moreover, the spin-coupling patterns suggested that the bonding was more akin to the Kekulé picture of benzene (with alternating double and single bonds) together with small contributions from Dewar benzene rather than the completely delocalised representation from molecular orbital theory.

3.7 Density Functional Theory

Density functional theory (DFT) is an approach to the electronic structure of atoms and molecules which has enjoyed a dramatic surge of interest since the late 1980s and 1990s [Parr 1983; Wimmer 1997]. Our approach here will be to introduce the key elements of the theory and to identify the similarities and differences between DFT and the Hartree–Fock approach. In Hartree–Fock theory the multi-electron wavefunction is expressed as a Slater determinant which is constructed from a set of N single-electron wavefunctions (N being the number of electrons in the molecule). DFT also considers single-electron functions. However, whereas Hartree–Fock theory does indeed calculate the full N-electron wavefunction, density functional theory only attempts to calculate the total electronic energy and the overall electronic density distribution. The central idea underpinning DFT is that

there is a relationship between the total electronic energy and the overall electronic density. This is not a particularly new idea; indeed an approximate model developed in the late 1920s (the Thomas–Fermi model) contains some of the basic elements. However, the real breakthrough came with a paper by Hohenberg and Kohn in 1964 [Hohenberg and Kohn 1964], who showed that the ground-state energy and other properties of a system were uniquely defined by the electron density. This is sometimes expressed by stating that the energy, E, is a unique *functional* of $\rho(\mathbf{r})$. A functional enables a function to be mapped to a number and is usually written using square brackets. Thus:

$$Q[f(\mathbf{r})] = \int f(\mathbf{r})\, d\mathbf{r} \tag{3.36}$$

The function $f(\mathbf{r})$ is usually dependent upon other well-defined functions. A simple example of a functional would be the area under a curve, which takes a function $f(\mathbf{r})$ defining the curve between two points and returns a number (the area, in this case). In the case of DFT the function depends upon the electron density, which would make Q a functional of $\rho(\mathbf{r})$, in the simplest case $f(\mathbf{r})$ would be equivalent to the density (i.e. $f(\mathbf{r}) \equiv \rho(\mathbf{r})$). If the function $f(\mathbf{r})$ were to depend in some way upon the gradients (or higher derivatives) of $\rho(\mathbf{r})$ then the functional is referred to as being 'non-local', or 'gradient-corrected'. By contrast, a 'local' functional would only have a simple dependence upon $\rho(\mathbf{r})$. In DFT the energy functional is written as a sum of two terms:

$$E[\rho(\mathbf{r})] = \int V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r})\, d\mathbf{r} + F[\rho(\mathbf{r})] \tag{3.37}$$

The first term arises from the interaction of the electrons with an external potential $V_{\text{ext}}(\mathbf{r})$ (typically due to the Coulomb interaction with the nuclei). $F[\rho(\mathbf{r})]$ is the sum of the kinetic energy of the electrons and the contribution from interelectronic interactions. The minimum value in the energy corresponds to the exact ground-state electron density, so enabling a variational approach to be used (i.e. the 'best' solution corresponds to the minimum of energy and an incorrect density gives an energy above the true energy). There is a constraint on the electron density as the number of electrons (N) is fixed:

$$N = \int \rho(\mathbf{r})\, d\mathbf{r} \tag{3.38}$$

In order to minimise the energy we introduce this constraint as a Lagrangian multiplier ($-\mu$), leading to:

$$\frac{\delta}{\delta\rho(\mathbf{r})}\left[E[\rho(\mathbf{r})] - \mu\int \rho(\mathbf{r})\, d\mathbf{r}\right] = 0 \tag{3.39}$$

From this we can write:

$$\left(\frac{\delta E[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})}\right)_{V_{\text{ext}}} = \mu \tag{3.40}$$

Equation (3.40) is the DFT equivalent of the Schrodinger equation. The subscript V_{ext} indicates that this is under conditions of constant external potential (i.e. fixed nuclear positions) It is interesting to note that the Lagrange multiplier, μ, can be identified with the chemical potential of an electron cloud for its nuclei, which in turn is related to the

electronegativity, χ:

$$-\chi = \mu = \left(\frac{\partial E}{\partial N}\right)_{V_{ext}} \tag{3.41}$$

The second landmark paper in the development of density functional theory was by Kohn* and Sham who suggested a practical way to solve the Hohnberg–Kohn theorem for a set of interacting electrons [Kohn and Sham 1965]. The difficulty with Equation (3.37) is that we do not know what the function $F[\rho(\mathbf{r})]$ is. Kohn and Sham suggested that $F[\rho(\mathbf{r})]$ should be approximated as the sum of three terms:

$$F[\rho(\mathbf{r})] = E_{KE}[\rho(\mathbf{r})] + E_H[\rho(\mathbf{r})] + E_{XC}[\rho(\mathbf{r})] \tag{3.42}$$

where $E_{KE}[\rho(\mathrm{r})]$ is the kinetic energy, $E_H[\rho(\mathrm{r})]$ is the electron–electron Coulombic energy, and $E_{XC}[\rho(\mathrm{r})]$ contains contributions from exchange and correlation. It is important to note that the first term in Equation (3.42), $E_{KE}[\rho(\mathrm{r})]$, is defined as the kinetic energy of a system of *non-interacting* electrons with the same density $\rho(\mathbf{r})$ as the real system:

$$E_{KE}[\rho(\mathrm{r})] = \sum_{i=1}^{N} \int \psi_i(\mathrm{r}) \left(-\frac{\nabla^2}{2}\right) \psi_i(\mathrm{r})\, d\mathbf{r} \tag{3.43}$$

The second term, $E_H(\rho)$, is also known as the Hartree electrostatic energy. The Hartree approach to solving the Schrodinger equation was introduced briefly in Section 2.3.3 and almost immediately dismissed because it fails to recognise that electronic motions are correlated. In the Hartree approach this electrostatic energy arises from the classical interaction between two charge densities, which, when summed over all possible pairwise interactions, gives:

$$E_H[\rho(\mathrm{r})] = \frac{1}{2} \iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|r_1 - r_2|}\, d\mathbf{r}_1\, d\mathbf{r}_2 \tag{3.44}$$

Combining these two and adding the electron–nuclear interaction leads to the full expression for the energy of an N-electron system within the Kohn–Sham scheme:

$$E[\rho(\mathbf{r})] = \sum_{i=1}^{N} \int \psi_i(\mathrm{r}) \left(-\frac{\nabla^2}{2}\right) \psi_i(\mathrm{r})\, d\mathbf{r} + \frac{1}{2} \iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\, d\mathbf{r}_1\, d\mathbf{r}_2 + E_{XC}[\rho(\mathbf{r})]$$
$$- \sum_{A=1}^{M} \int \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \rho(\mathrm{r})\, d\mathbf{r} \tag{3.45}$$

This equation acts to *define* the exchange-correlation energy functional $E_{XC}[\rho(\mathrm{r})]$, which thus contains not only contributions due to exchange and correlation but also a contribution due to the difference between the true kinetic energy of the system and $E_{KE}[\rho(\mathbf{r})]$.

* Walter Kohn, whose name appears on the two key papers which provided the impetus for the development of 'modern' density functional theory, was awarded the Nobel Prize for Chemistry in 1998, jointly with John Pople

Kohn and Sham wrote the density $\rho(\mathbf{r})$ of the system as the sum of the square moduli of a set of one-electron orthonormal orbitals:

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} |\psi_i(\mathbf{r})|^2 \tag{3 46}$$

By introducing this expression for the electron density and applying the appropriate variational condition the following one-electron Kohn–Sham equations result:

$$\left\{-\frac{\nabla_1^2}{2} - \left(\sum_{A=1}^{M} \frac{Z_A}{r_{1A}}\right) + \int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + V_{XC}[\mathbf{r}_1]\right\}\psi_i(\mathbf{r}_1) = \varepsilon_i \psi_i(\mathbf{r}_1) \tag{3.47}$$

In Equation (3.47) we have written the external potential in the form appropriate to the interaction with M nuclei. ε_i are the orbital energies and V_{XC} is known as the exchange-correlation functional, related to the exchange-correlation energy by:

$$V_{XC}[\mathbf{r}] = \left(\frac{\delta E_{XC}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})}\right) \tag{3.48}$$

The total electronic energy is then calculated from Equation (3.45).

To solve the Kohn–Sham equations a self-consistent approach is taken. An initial guess of the density is fed into Equation (3.47) from which a set of orbitals can be derived, leading to an improved value for the density, which is then used in the second iteration, and so on until convergence is achieved.

3.7.1 Spin-polarised Density Functional Theory

Local spin density functional theory (LSDFT) is an extension of 'regular' DFT in the same way that restricted and unrestricted Hartree–Fock extensions were developed to deal with systems containing unpaired electrons. In this theory both the electron density and the spin density are fundamental quantities with the net spin density being the difference between the density of up-spin and down-spin electrons:

$$\sigma(\mathbf{r}) = \rho_\uparrow(\mathbf{r}) - \rho_\downarrow(\mathbf{r}) \tag{3.49}$$

The total electron density is just the sum of the densities for the two types of electron. The exchange-correlation functional is typically different for the two cases, leading to a set of spin-polarised Kohn–Sham equations:

$$\left\{-\frac{\nabla_1^2}{2} - \left(\sum_{A=1}^{M} \frac{Z_A}{r_{1A}}\right) + \int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + V_{XC}[\mathbf{r}_1, \sigma]\right\}\psi_i^\sigma(\mathbf{r}_1) = \varepsilon_i^\sigma \psi_i^\sigma(\mathbf{r}_1) \quad \sigma = \alpha, \beta \tag{3.50}$$

This leads to two sets of wavefunctions, one for each spin, similar to UHF theory.

3.7.2 The Exchange-correlation Functional

The exchange-correlation functional is clearly key to the success (or otherwise) of the density functional approach. One reason why DFT is so appealing is that even relatively simple

approximations to the exchange-correlation functional can give favourable results. The simplest way to obtain this contribution uses the so-called *local density approximation* (LDA; the acronym LSDA is also used, for local spin density approximation), which is based upon a model called the uniform electron gas, in which the electron density is constant throughout all space. The total exchange-correlation energy, E_{XC}, for our system can then be obtained by integrating over all space:

$$E_{XC}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\varepsilon_{XC}(\rho(\mathbf{r}))\, d\mathbf{r} \tag{3.51}$$

$\varepsilon_{XC}(\rho(\mathbf{r}))$ is the exchange-correlation energy per electron as a function of the density in the uniform electron gas. The exchange-correlation functional is obtained by differentiation of this expression:

$$V_{XC}[\mathrm{r}] = \rho(\mathbf{r})\frac{d\varepsilon_{XC}(\rho(\mathbf{r}))}{d\rho(\mathbf{r})} + \varepsilon_{XC}(\rho(\mathbf{r})) \tag{3.52}$$

In the local density approximation it is assumed that at each point **r** in the inhomogeneous electron distribution (i.e. in the system of interest) where the density is $\rho(\mathbf{r})$ then $V_{XC}[\rho(\mathbf{r})]$ and $\varepsilon_{XC}(\rho(\mathbf{r}))$ have the same values as in the homogeneous electron gas. In other words, the real electron density surrounding a volume element at position **r** is replaced by a constant electron density with the same value as at **r**. However, this 'constant' electron density is different for each point in space (Figure 3.6).

The exchange-correlation energy per electron (i.e. the energy density) of the uniform electron gas is known accurately for all densities of practical interest from various approaches such as quantum Monte Carlo methods [Ceperley and Alder 1980]. In order to be of practical use this exchange-correlation energy density is then expressed in an analytical form that makes it amenable to computation. It is usual to express $\varepsilon_{XC}[\rho(\mathbf{r})]$ as an analytical function of the electron density and to consider the exchange and correlation contributions separately. However, some analytical expressions for the combined exchange and correlation energy density do exist, such as the following expression of Gunnarsson and

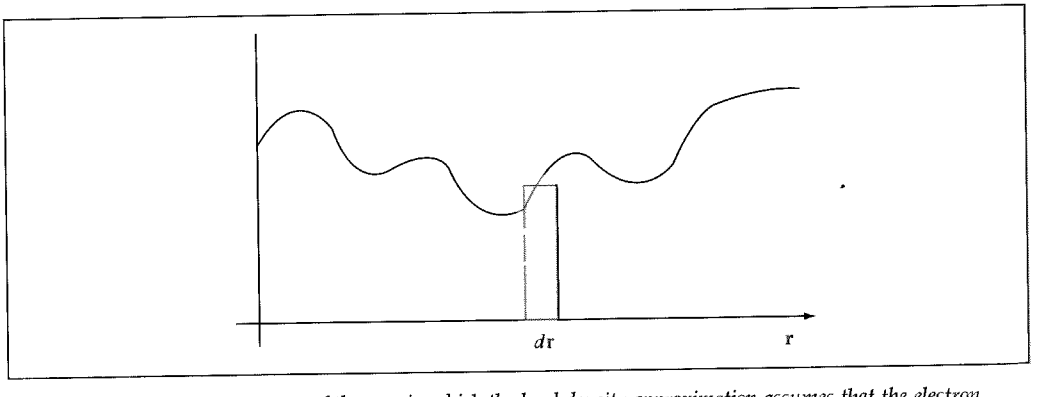

Fig 3 6 Schematic representation of the way in which the local density approximation assumes that the electron density within a volume element dr surrounding a point r is assumed to be constant.

Lundqvist [Gunnarsson and Lundqvist 1976]:

$$\varepsilon_{XC}(\rho(\mathbf{r})) = -\frac{0.458}{r_s} - 0.0666G\left(\frac{r_s}{11.4}\right);$$

$$G(x) = \frac{1}{2}\left[(1+x)\log(1+x^{-1}) - x^2 + \frac{x}{2} - \frac{1}{3}\right], \quad r_s^3 = \frac{3}{4\pi\rho(\mathbf{r})} \tag{3.53}$$

The following relatively simple expression is commonly used for the exchange-only energy under the local density approximation [Slater 1974]:

$$E_X[\rho_\alpha(\mathbf{r}), \rho_\beta(\mathbf{r})] = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3} \int (\rho_\alpha^{4/3}(\mathbf{r}) + \rho_\beta^{4/3}(\mathbf{r}))\, d\mathbf{r} \tag{3.54}$$

where α and β represent up and down spins. In general, more attention has been paid to the correlation contribution, for which there is no such simple functional form. Perdew and Zunger suggested the following parametric relationship for the correlation contribution [Perdew and Zunger 1981]:

$$\varepsilon_C(\rho(\mathbf{r})) = \begin{cases} -0.1423/(1 + 1.9529r_s^{1/2} + 0.3334r_s) & r_s \geq 1 \\ -0.0480 + 0.0311\ln r_s - 0.0116r_s + 0.0020r_s \ln r_s & r_s < 1 \end{cases} \tag{3.55}$$

This result applies when the number of up spins equals the number of down spins and so is not applicable to systems with an odd number of electrons. The correlation energy functional was also considered by Vosko, Wilk and Nusair [Vosko *et al.* 1980], whose expression is:

$$\varepsilon_C(\rho(\mathbf{r})) = \frac{A}{2}\left\{\ln\frac{x^2}{X(x)} + \frac{2b}{Q}\tan^{-1}\frac{Q}{2x+b} - \frac{bx_0}{X(x_0)}\left[\ln\frac{(x-x_0)^2}{X(x)} + \frac{2(b+2x_0)}{Q}\tan^{-1}\frac{Q}{2x+b}\right]\right\}$$

$$x = r_s^{1/2}, \quad X(x) = x^2 + bx + c, \quad Q = (4c - b^2)^{1/2}; \tag{3.56}$$

$$A = 0.062\,1814, \quad x_0 = -0.409\,286, \quad b = 13.0720, \quad c = 42.7198$$

In addition to the energy terms for the exchange-correlation contribution (which enables the total energy to be determined) it is necessary to have corresponding terms for the potential, $V_{XC}[\rho(\mathbf{r})]$, which are used to solve the Kohn–Sham equations. These are obtained as the appropriate first derivatives using Equation (3.52).

To solve the Kohn–Sham equations a number of different approaches and strategies have been proposed. One important way in which these can differ is in the choice of basis set for expanding the Kohn–Sham orbitals. In most (but not all) DFT programs for calculating the properties of molecular systems (rather than for solid-state materials) the Kohn–Sham orbitals are expressed as a linear combination of atomic-centred basis functions:

$$\psi_i(\mathbf{r}) = \sum_{\nu=1}^{K} c_{\nu i}\phi_\nu \tag{3.57}$$

Several functional forms have been investigated for the basis functions ϕ_ν. Given the vast experience of using Gaussian functions in Hartree–Fock theory it will come as no surprise to learn that such functions have also been employed in density functional theory. However, these are not the only possibility: Slater type orbitals are also used, as are *numerical*

basis functions. We encountered Slater type orbitals in Chapter 2, but the notion of a numerical basis function is new. A numerical basis function can be generated by solving the Kohn–Sham equations for isolated atoms. This gives a set of values on a spherical polar grid centred on each atom. The variation at each grid point can be stored as a cubic spline function so enabling analytical gradients to be calculated. One advantage of a numerical basis set (if properly derived) is that it has the correct nodal behaviour close to the nucleus together with an exponential decay.

More than one function may be used to represent a particular atomic orbital. This is obviously a well-understood tactic when using Gaussian functions, but the use of basis set contractions also applies to the Slater type orbitals and the numerical basis sets. For a numerical basis set the 'contraction' can be derived from two functions, one corresponding to the neutral atom and the other to a positive ion.

If the basis set expansion for the Kohn–Sham orbitals in Equation (3.57) is substituted into the Kohn–Sham equations then it is possible to express them in a matrix form, identical in form to the Roothaan–Hall equations:

$$\mathbf{HC} = \mathbf{SCE} \tag{3.58}$$

In this matrix equation the elements of the Kohn–Sham matrix H are given by:

$$H_{\mu\nu} = \int d\mathbf{r}_1 \phi_\mu(\mathbf{r}_1) \left\{ -\frac{\nabla_1^2}{2} - \left(\sum_{A=1}^{M} \frac{Z_A}{r_{1A}} \right) + \int \frac{\rho(\mathbf{r}_2)}{r_{12}} \, d\mathbf{r}_2 + V_{\mathrm{XC}}[\mathbf{r}_1] \right\} \phi_\nu(\mathbf{r}_1) \tag{3.59}$$

The first two terms are straightforward and are equal to the core contribution, $H_{\mu\nu}^{\mathrm{core}}$. The Coulomb repulsion contribution (the Hartree term) can be expanded in terms of the basis functions and the density matrix, P:

$$\iint \frac{\phi_\mu(\mathbf{r}_1)\rho(\mathbf{r}_2)\phi_\nu(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|} \, d\mathbf{r}_1 \, d\mathbf{r}_2 = \sum_{\lambda=1}^{K} \sum_{\sigma=1}^{K} P_{\lambda\sigma} \iint \frac{\phi_\mu(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)\phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \, d\mathbf{r}_1 \, d\mathbf{r}_2 \tag{3.60}$$

For a closed-shell system with N electrons the elements of the density matrix are given by:

$$P_{\mu\nu} = 2 \sum_{i=1}^{N/2} c_{\mu i} c_{\nu i} \tag{3.61}$$

This is just the same as for the Roothaan-Hall approach to Hartree–Fock theory. The overlap matrix, S, is defined similarly:

$$S_{\mu\nu} = \int \phi_\mu(\mathbf{r}) \phi_\nu(\mathbf{r}) \, d\mathbf{r} \tag{3.62}$$

The overall procedure to achieve self-consistency is very reminiscent of that used in Hartree–Fock theory, involving first an initial guess of the density by superimposing atomic densities, construction of the Kohn–Sham and overlap matrices, and diagonalisation to give the eigenfunctions and eigenvectors from which the Kohn–Sham orbitals* can be

* It is important to note that the Kohn–Sham orbitals used in density functional theory are a set of non-interacting orbitals designed to give the correct density and have no physical meaning beyond that, unlike the orbitals used in Hartree–Fock theory

constructed and thus the density for the next iteration. This cycle continues until convergence is achieved.

The appearance of the four-centre integrals in Equation (3.60) might lead one to question the advantage of the DFT approach, at least as far as computational efficiency is concerned. Whilst these integrals can certainly be tackled using the same techniques as in Hartree–Fock theory, it is also viable in density functional theory to avoid having to calculate them by considering the left-hand side of Equation (3.60). There are two basic ways to do this. First, one can approximate the charge density by another basis set expansion:

$$\rho(\mathbf{r}) \approx \sum_k c_k \phi'_k(\mathbf{r}) \tag{3.63}$$

These auxiliary basis functions ϕ' have the same functional form as the orbital expansion and the coefficients c_k are obtained by a least-squares fitting procedure. Substituting for the density in the four-centre integrals gives a computationally less demanding three-centre, two-electron integral:

$$\iint \frac{\phi_\mu(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)\phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\, d\mathbf{r}_1\, d\mathbf{r}_2 = \iint \frac{\phi_\mu(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)\phi'_k(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\, d\mathbf{r}_1\, d\mathbf{r}_2 \tag{3.64}$$

The second approach focuses on the Coulomb integral and uses Poisson's equation. Let us introduce $V_{\mathrm{el}}(\mathbf{r}_1)$:

$$V_{\mathrm{el}}(\mathbf{r}_1) = \int \frac{\rho(r_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\, d\mathbf{r}_2 \tag{3.65}$$

Poisson's equation relates the second derivative of the electric potential to the charge density:

$$\nabla^2 V(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{3.66}$$

We can thus write:

$$\nabla^2 \int \frac{\rho(r_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\, d\mathbf{r}_2 = -4\pi\rho(\mathbf{r}_1) \tag{3.67}$$

This equation can be solved numerically on a grid to determine $V_{\mathrm{el}}(\mathbf{r}_1)$ The same grid is then used to numerically integrate the four-centre, two-electron integral, Equation (3.60), as follows:

$$\iint \frac{\phi_\mu(\mathbf{r}_1)\rho(\mathbf{r}_2)\phi_\nu(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|}\, d\mathbf{r}_1\, d\mathbf{r}_2 \equiv \int \phi_\mu(\mathbf{r}_1) V_{\mathrm{el}}(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1) \approx \sum_{i=1}^{P} \phi_\mu(\mathbf{R}_i) V_{\mathrm{el}}(\mathbf{R}_i)\phi_\nu(\mathbf{R}_i) W_i \tag{3.68}$$

In this equation the P points $\mathbf{R}_i$ correspond to the grid used to solve the Poisson equation for V_{el} and W_i are weighting factors.

It might be wondered why these two simplifications for the four-centre, two-electron integrals can be used in density functional theory and not in Hartree–Fock theory. The reason is that the exchange contribution in Hartree–Fock theory is not a function that can be simplified (technically, it is a non-local functional), in contrast to the situation in

density functional theory. As the four-centre integrals must therefore still be determined for the exchange component in Hartree–Fock theory there is nothing to be gained from simplifying the corresponding Coulomb term.

The exchange-correlation contribution to the Kohn–Sham matrix elements (the final term in Equation (3.59)) is invariably evaluated using a grid of points. This is a consequence of the complexity of the functionals employed. The integration may then be performed using the grid directly or by fitting a further auxiliary basis set expansion with which analytical integration can be used. If a DFT program uses a basis set containing K functions and employs either a grid-based integration scheme with P points or an auxiliary basis set with P functions then the computational complexity of the calculation scales as K^2P. As P is often linearly related to K, density functional theory is often said to scale as the cube of the number of basis functions, K^3. This contrasts with the fourth-power scaling for conventional Hartree–Fock calculations. However, many practical density functional calculations with a well-engineered computer program do not scale as the simple third power, just as practical Hartree–Fock calculations do not scale as the fourth power; these oft-quoted statements apply only to the most naïve implementations or for calculations on very small, test systems where integral neglect thresholds are not employed.

Whilst most of the programs which use density functional theory for molecular calculations employ one of the three types of basis set described thus far, there are two important alternatives to this approach. The first of these involves the solution of the Kohn–Sham equations numerically (on a grid) using what is sometimes referred to as a 'basis-set free' approach [Becke and Dickson 1990]. Such an approach is thus free from the limitations of a finite basis set expansion (provided, of course, that sufficient grid points are employed!) and can be used to evaluate different exchange-correlation functionals, as these represent the only remaining source of error. The second alternative is particularly important for the study of bulk systems such as metals and alloys and involves the use of *plane waves*. This approach will be discussed later in this chapter when we consider the general problem of using quantum mechanics to study the solid state.

3.7.3 Beyond the Local Density Approximation: Gradient-corrected Functionals

The most important feature of density functional theory is probably the way in which it directly incorporates exchange and correlation effects; the latter in particular are only truly considered in the more complex, post-Hartree–Fock approaches such as configuration interaction or many-body perturbation theory. Despite its simplicity the local density approximation performs surprisingly well. However, the local density approximation has been shown to be clearly inadequate for some problems and for this reason extensions have been developed. The most common method is to use gradient-corrected, 'non-local' functionals which depend upon the gradient of the density at each point in space and not just on its value. These gradient corrections are typically divided into separate exchange and correlation contributions. A variety of gradient corrections have been proposed in the literature. The gradient correction to the exchange functional proposed by Becke is popular [Becke 1988, 1992]; this corrects

the local spin density approximation result as follows:

$$E_X[\rho(\mathbf{r})] = E_X^{LSDA}[\rho(\mathbf{r})] - b\sum_{\sigma=\alpha,\beta}\int \rho_\sigma^{4/3}\frac{x_\sigma^2}{(1+6bx_\sigma \sinh^{-1}x_\sigma)}d\mathbf{r}; \quad x_\sigma = \frac{|\nabla\rho_\sigma|}{\rho_\sigma^{4/3}} \tag{3.69}$$

$E_X^{LSDA}[\rho(\mathbf{r})]$ is the standard Slater form of the exchange energy, Equation (3.54). The form written in Equation (3.69) is for a spin-unrestricted system, from which the appropriate expression for a closed-shell system is easily derived. x_σ is a dimensionless parameter and b is constant with a value of 0.0042 a.u. The value of b was determined by fitting to exact exchange Hartree–Fock energies for the noble gas atoms helium to radon. Two particular features of this functional form are that in the limit $r \to \infty$ the limiting form of the exchange-correlation integral is correctly achieved and that it uses just a single parameter, b. The correlation functional of Lee, Yang and Parr is also widely used [Lee *et al.* 1988]; in its original form it was expressed as follows (for a closed-shell system):

$$E_C[\rho(\mathbf{r})] = -a\int \frac{1}{1+d\rho^{-1/3}}\{r + b\rho^{-2/3}[C_F\rho^{5/3} - 2t_W + (\tfrac{1}{9}t_W + \tfrac{1}{18}\nabla^2\rho)e^{-cr^{-1/3}}]\}\,d\mathbf{r}$$
$$t_W(\mathbf{r}) = \sum_{i=1}^{N}\frac{|\nabla\rho_i(\mathbf{r})|^2}{\rho_i(\mathbf{r})} - \tfrac{1}{8}\nabla^2\rho; \quad C_F = \tfrac{3}{10}(3\pi^2)^{2/3} \tag{3.70}$$

a, b, c and d are constants with values 0.049, 0.132, 0.2533 and 0.349, respectively. This expression provides both local and non-local components within a single expression and the gradient contribution to second order. A combination of the standard local spin density approximation exchange result (Equation (3.54)) with the Becke gradient-exchange correction and the Lee–Yang–Parr correlation functional is currently a popular choice, commonly abbreviated to BLYP (pronounced 'blip').

3.7.4 Hybrid Hartree–Fock/Density Functional Methods

As we stated earlier, a key feature of density functional theory is the way in which correlation effects are incorporated from the beginning, unlike Hartree–Fock theory. Moreover, the incorporation of correlation into the Hartree–Fock formalism often involves significant computational overhead, as we have considered in Section 3.3. However, it is important to recognise that Hartree–Fock theory does provide an essentially exact means of treating the exchange contribution. One potentially attractive option is thus to add a correlation energy derived from DFT (e.g. the local density approximation) to the Hartree–Fock energy. In such an approach the exchange-correlation energy is written as a sum of the exact exchange term together with the correlation component from the local density approximation. This 'exact' exchange energy is obtained from the Slater determinant of the Kohn–Sham orbitals.

Unfortunately, this simple approach does not work well, but Becke has proposed a strategy which does seem to have much promise [Becke 1993a, b]. In his approach the exchange-correlation energy E_{XC} is written in the following form:

$$E_{XC} = \int_0^1 U_{XC}^{\lambda}\,d\lambda \tag{3.71}$$

Equation (3.71) contains a coupling parameter λ, which takes values from 0 to 1. A value of zero corresponds to a system where there is no Coulomb repulsion U_{XC} between the electrons (i.e. the Kohn–Sham non-interacting reference state). As λ increases to 1 the interelectronic Coulomb repulsion is introduced until $\lambda = 1$, which corresponds to the 'real' system with full interactions. For all values of λ the electron density is the same and equal to the density of the real system. It is not practical to perform this integral analytically and so it must be approximated. The simplest approximation is a linear interpolation:

$$E_{XC} = \tfrac{1}{2}\left(U_{XC}^{0} + U_{XC}^{1}\right) \tag{3.72}$$

When $\lambda = 0$ we have U_{XC}^{0}, which is the exchange-correlation potential energy of the non-interacting reference system As there are no electronic interactions in this system there is no correlation term and so U_{XC}^{0} corresponds to the pure exchange energy of the Kohn–Sham determinant and can be determined exactly. U_{XC}^{1} is the exchange-correlation potential energy of the full-interacting real system. Becke proposed that this should be calculated using the local spin-density approximation. This potential energy (note that it is not the total energy, E) is available from:

$$U_{XC}^{1} \approx U_{XC}^{LSDA} = \int u_{XC}\left[\rho_{\alpha}(\mathbf{r}), \rho_{\beta}(\mathbf{r})\right] d\mathbf{r} \tag{3.73}$$

u_{XC} is the exchange-correlation potential energy density of an electron gas for which appropriate expressions are available.

This so-called 'half-and-half' theory proved to be significantly better than the alternative methods based upon mixing exact exchange and correlation energies. In a refinement of the scheme, Becke recognised that there were problems with the model when $\lambda = 0$. These problems arise because the electron gas model is not appropriate near this exchange-only limit for molecular bonds. Hence a key feature of Becke's modified model is to eliminate the term U_{XC}^{0} and to write the exchange-correlation energy as the following linear combination:

$$E_{XC} = E_{XC}^{LSDA} + a_0\left(E_X^{exact} - E_X^{LSDA}\right) + a_X \Delta E_X^{GC} + a_C \Delta E_C^{GC} \tag{3.74}$$

In Equation (3.74) E_X^{exact} is the exact exchange energy (obtained from the Slater determinant of the Kohn–Sham orbitals), E_X^{LSDA} is the exchange energy under the local spin density approximation, ΔE_X^{GC} is the gradient correction for exchange and ΔE_C^{GC} is the gradient correction for correlation. a_0, a_X and a_C are empirical coefficients obtained by least-squares fitting to experimental data (56 atomisation energies, 42 ionisation potentials, eight proton affinities and the total atomic energies of the ten first-row elements). Their values are $a_0 = 0.20$, $a_X = 0.72$ and $a_C = 0.81$. In Becke's original paper his own gradient correction for exchange was used together with a gradient correction for correlation suggested by Perdew and Wang. An alternative to this scheme is to employ the Lee–Yang–Parr correlation functional (with the gradient term) and the standard local correlation functional due to Vosko, Wilk and Nusair (VWN). This is the 'B3LYP' density functional:

$$E_{XC}^{B3LYP} = (1 - a_0)E_X^{LSDA} + a_0 E_X^{HF} + a_X \Delta E_X^{B88} + a_C E_C^{LYP} + (1 - a_C)E_C^{VWN} \tag{3.75}$$

3.7.5 Performance and Applications of Density Functional Theory

The application of density functional theory to isolated, 'organic' molecules is still in relative infancy compared with the use of Hartree–Fock methods. There continues to be a steady stream of publications designed to assess the performance of the various approaches to DFT As we have discussed there is a plethora of ways in which density functional theory can be implemented with different functional forms for the basis set (Gaussians, Slater type orbitals, or numerical), different expressions for the exchange and correlation contributions within the local density approximation, different expressions for the gradient corrections and different ways to solve the Kohn–Sham equations to achieve self-consistency. This contrasts with the situation for Hartree–Fock calculations, which mostly use one of a series of tried and tested Gaussian basis sets and where there is a substantial body of literature to help choose the most appropriate method for incorporating post-Hartree–Fock methods, should that be desired.

A clear conclusion from such comparative studies is that density functional methods using gradient-corrected functionals can give results for a wide variety of properties that are competitive with, and in some cases superior to, *ab initio* calculations using correlation (e.g. MP2). Gradient-corrected functionals are required for the calculation of relative conformational energies and the study of intermolecular systems, particularly those involving hydrogen bonding [Sim *et al.* 1992]. As is the case with the *ab initio* methods the choice of basis set is also important in determining the results. By keeping the basis set constant (6-31G* being a popular choice) it is possible to make objective comparisons. Four examples of such comparative studies are those of Johnson and colleagues, who considered small neutral molecules [Johnson *et al.* 1993]; St-Amant *et al.*, who examined small organic molecules [St-Amant *et al.* 1995]; Stephens *et al.*, who performed a detailed study of the absorption and circular dichroism spectra of 4-methyl-2-oxetanone [Stephens *et al.* 1994]; and Frisch *et al.*, who compared a variety of density functional methods with one another and to traditional *ab initio* approaches [Frisch *et al.* 1996]. The evolution of defined sets of data such as those associated with the Gaussian-*n* series of models has also acted as a spur to those involved in developing density functional methods. For example, much of Becke's work on gradient corrections and on mixed Hartree–Fock/density function methods was evaluated using data sets originally collated for the Gaussian-1 and Gaussian-2 methods. A more recent example is a variant of the Gaussian-3 method which uses B3LYP to determine geometries and zero-point energies [Baboul *et al.* 1999].

One of the most important developments for the practical application of DFT were methods for calculating analytical gradients of the energy with respect to the nuclear coordinates. This enables molecular geometries to be optimised. Once more there are some differences between the way this is done with density functional theory compared with Gaussian-based Hartree–Fock methods. A potential problem is that the use of grid-based integration schemes makes it difficult to provide exact expressions for the gradients. However, the errors associated with the grid-based method are generally very small and do not cause problems during the optimisation.

3.8 Quantum Mechanical Methods for Studying the Solid State

3.8.1 Introduction

The quantum mechanical methods used to study the behaviour of solid-phase systems are often somewhat different to those traditionally employed for studies of individual molecules or isolated intermolecular complexes. A perfect crystalline system can be constructed by stacking copies of some repeating unit (the *unit cell*) in a systematic fashion without overlapping and without gaps. The structure of a crystal can be specified by defining the size and shape of the unit cell and the positions of the atoms within it. The unit cell is parallelepiped in shape and is characterised by three lattice vectors **a**, **b** and **c** and the angles between them (Figure 3.7). It may be possible to conceive of more than one unit cell, with different unit cell parameters In such cases a set of standard cell parameters can be obtained by the application of standardisation rules. The coordinates of the atoms in the unit cell may be expressed as fractional coordinates ($\alpha\mathbf{a}$, $\beta\mathbf{b}$, $\gamma\mathbf{c}$). Indeed, any general vector **r** can be written in terms of these basis vectors:

$$\mathbf{r} = (\alpha\mathbf{a}, \beta\mathbf{b}, \gamma\mathbf{c}) \tag{3 76}$$

where α, β and γ are not necessarily restricted to values between 0 and 1. There are fourteen different types of basic unit cell; these are the *Bravais lattices*. Common Bravais lattices include the simple cubic, body-centred cubic and face-centred cubic (Figure 3.8). Another common structure also shown in Figure 3.8 is the hexagonal close-packed arrangement, for which the underlying Bravais lattice (called the simple hexagonal) is formed from an underlying triangular arrangement. In addition to the translational symmetry that the unit cell must possess there may be some symmetry to the arrangement of the atoms within the unit cell. The particular combination of symmetry elements in a crystal defines its *space group*. There are 230 different space groups. If there is symmetry within the unit cell then it is strictly only necessary to specify the asymmetric unit (the unique part of the structure); the positions of the other atoms can be generated using the appropriate symmetry operators.

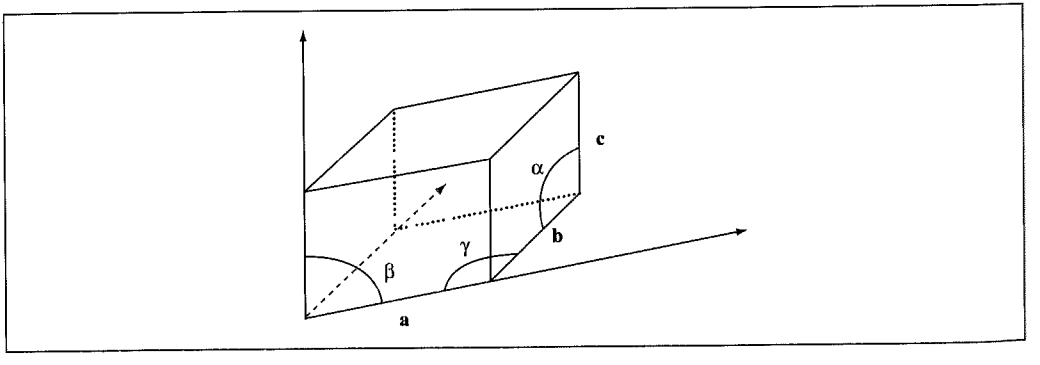

*Fig. 3 7· The six parameters **a**, **b**, **c**, α, β, γ which characterise the unit cell.*

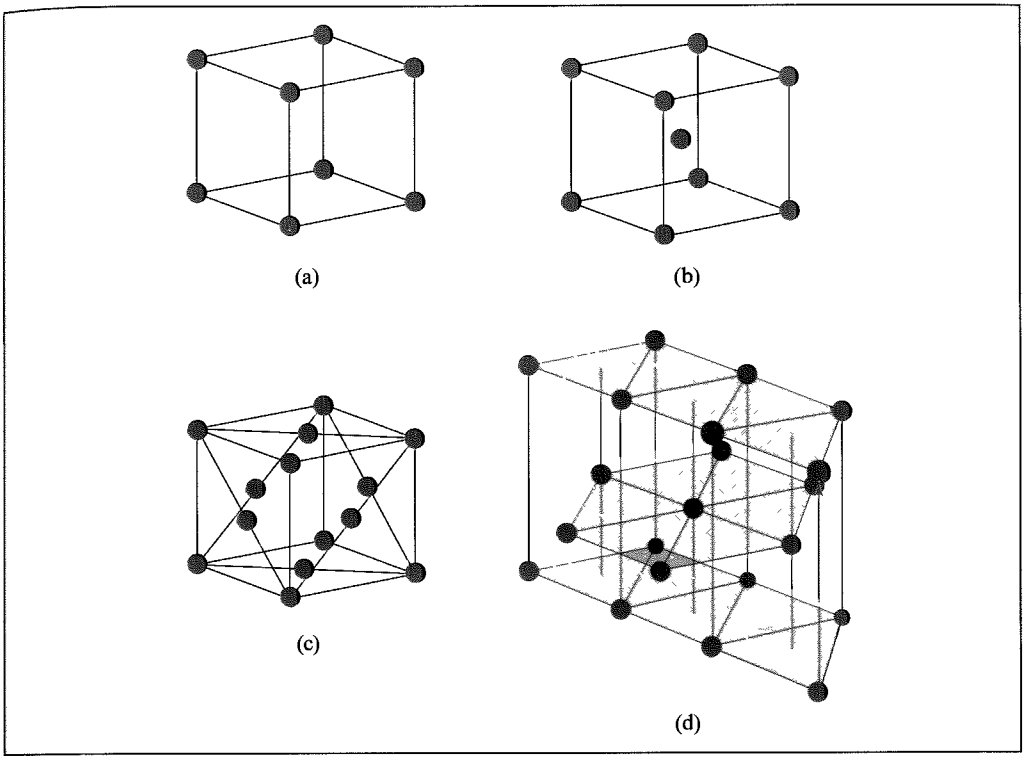

Fig. 3 8 Some basic Bravais lattices (a) simple cubic, (b) body-centred cubic, (c) face-centred cubic and (d) simple hexagonal close-packed (Figure adapted in part from Ashcroft N W and Mermin N D 1976. Solid State Physics *New York, Holt, Rinehart and Winston.)*

Another concept that is extremely powerful when considering lattice structures is the *reciprocal lattice*. X-ray crystallographers use a reciprocal lattice defined by three vectors $\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$ in which $\mathbf{a}^*$ is perpendicular to $\mathbf{b}$ and $\mathbf{c}$ and is scaled so that the scalar product of $\mathbf{a}^*$ and $\mathbf{a}$ equals 1. $\mathbf{b}^*$ and $\mathbf{c}^*$ are similarly defined. In three dimensions this leads to the following definitions:

$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}}; \quad \mathbf{b}^* = \frac{\mathbf{a} \times \mathbf{c}}{\mathbf{b} \cdot \mathbf{a} \times \mathbf{c}}; \quad \mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{\mathbf{c} \cdot \mathbf{a} \times \mathbf{b}} \tag{3.77}$$

Note that the denominator in each case is equal to the volume of the unit cell. The fact that $\mathbf{a}^*$, $\mathbf{b}^*$ and $\mathbf{c}^*$ have the units of 1/length gives rise to the terms 'reciprocal space' and 'reciprocal lattice'. It turns out to be convenient for our computations to work with an expanded reciprocal space that is defined by three closely related vectors $\mathbf{a}^\$$, $\mathbf{b}^\$$ and $\mathbf{c}^\$$, which are multiples by 2π of the X-ray crystallographic reciprocal lattice vectors:

$$\mathbf{a}^\$ = 2\pi\mathbf{a}^*; \quad \mathbf{b}^\$ = 2\pi\mathbf{b}^*; \quad \mathbf{c}^\$ = 2\pi\mathbf{c}^* \tag{3.78}$$

A simple illustrative example of reciprocal space is that of a 2D square lattice where the vectors $\mathbf{a}$ and $\mathbf{b}$ are orthogonal and of length equal to the lattice spacing, a. Here $\mathbf{a}^*$ and $\mathbf{b}^*$ are directed along the same directions as $\mathbf{a}$ and $\mathbf{b}$ respectively and have a length $1/a$

combined to give the equivalent of molecular orbitals. It is based on the assumption that the effect of orbital overlap is to modulate but not change completely the initial atomic levels. The approximation is traditionally considered most useful for describing the electronic structure of systems such as insulators and transition metals with partially filled d shells. The second approach is called the *nearly free-electron approximation*. This theory starts by considering the electrons as free particles whose motion is modulated by the presence of the lattice. The nearly free-electron approximation is traditionally considered the more suitable approach to systems such as metals where there is substantial overlap of the valence orbitals. We will outline both approaches in turn, making use of some of the fundamental principles and properties of lattices discussed earlier.

3.8.2 Band Theory and Orbital-based Approaches

Band theory is perhaps easier for chemists to understand, starting as it does from an orbital picture. We will therefore spend somewhat less space discussing this than the nearly free-electron approximation. We will start by considering the simplest problem, a 1D lattice. Initially we consider what happens if we bring together two atoms along the x axis until they are separated by a distance, a. If each atom has one s orbital, then the combined system has two molecular orbitals (one bonding and one anti-bonding). If we then add a third atom then three molecular orbitals are obtained (one bonding, one non-bonding and one anti-bonding). Four atoms give four energy levels, and so on. As more atoms are added the energy levels merge to give what is an essentially continuous *band* of energy levels (Figure 3.11). Each energy level can accommodate two electrons so if each atom contributes one electron the band will be half full. The presence of unoccupied energy levels near to the top of the filled level means that it is very easy to excite electrons from the filled to the unfilled levels. The electrons are consequently very mobile, giving rise to the special conduction and thermal properties of a metal. By contrast, if each atom contributes two electrons then the band will be completely filled. Such electrons would have to be excited to higher bands, which might, for example, be formed by the overlap of p orbitals. In an insulator the energy of this p band would typically be significantly higher than the lower s band and so excitation would require considerable energy. In a semiconductor the band gap is smaller and it may be possible to excite electrons from the top of the highest filled band (the *valence band*) to the lowest unoccupied band (the *conduction band*) at normal temperatures. These three difference scenarios are illustrated in Figure 3.11.

The periodicity of the lattice means that the values of a function (such as the electron density) will be identical at equivalent points on the lattice. Likewise there is a relationship between the wavefunction at a point (x in our 1D lattice) and at an equivalent point elsewhere on the lattice (for the 1D lattice this would be $x + na$, where n is an integer). *Bloch's theorem* provides the link; each allowed lattice wavefunction must satisfy the following relationship:

$$\psi^k(x + a) = e^{ika}\psi^k(x) \tag{3.81}$$

In this equation we have identified the wavefunction with a label, k, which for now can be considered an index; there are as many values of k as there are atoms in the 1D lattice.

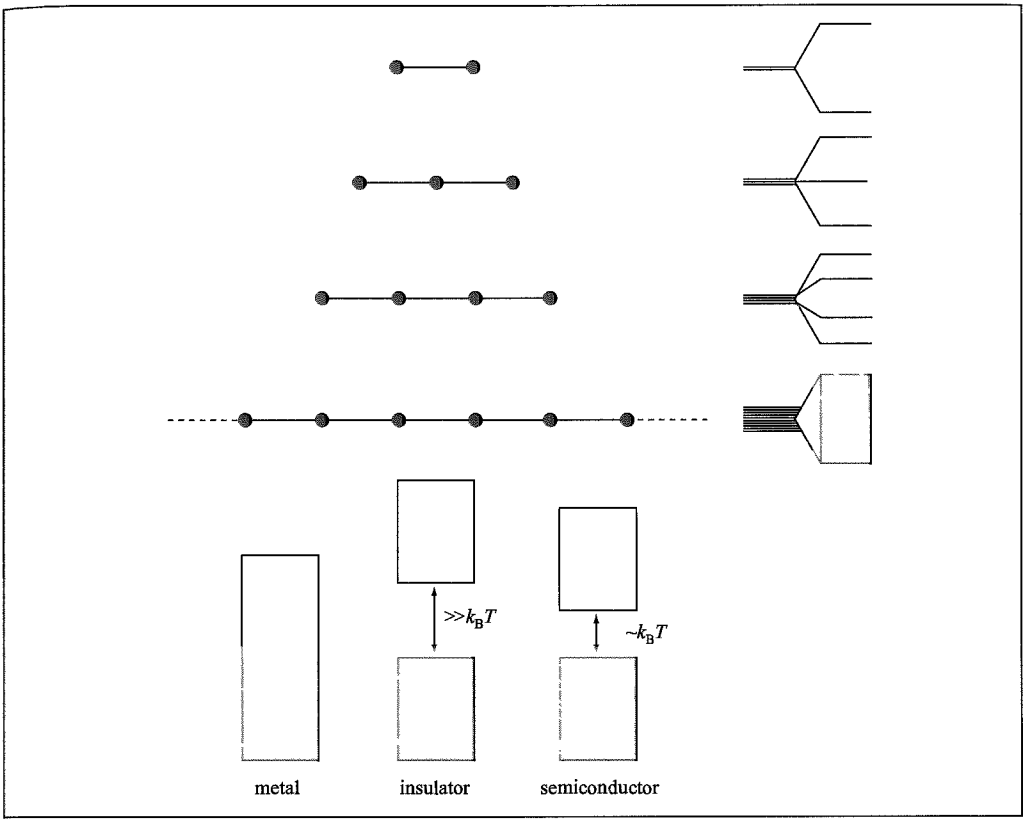

Fig 3 11: The creation of a band of energy levels from the overlap of two, three, four, etc atomic orbitals, which eventually gives rise to a continuum Also shown are the conceptual differences between metals, insulators and semiconductors

We wish to construct linear combinations of the atomic orbitals such that the overall wavefunction meets the Bloch requirement. Suppose the s orbitals in our lattice are labelled χ_n, where the nth orbital is located at position $x = na$. An acceptable linear combination of these orbitals that satisfies the Bloch requirements is:

$$\psi^k = \sum_n e^{ikna} \chi_n \tag{3.82}$$

We now need to consider how the form of the wavefunction varies with k. The first situation we consider corresponds to $k = 0$, where the exponential terms are all equal to 1 and the overall wavefunction becomes a simple additive linear combination of the atomic orbitals:

$$\psi^{k=0} = \sum_n \chi_n = \chi_0 + \chi_1 + \chi_2 + \cdots \tag{3.83}$$

The other situation we consider is $k = \pi/a$. Recall that $\exp(ix)$ can be written $\cos(x) + i\sin(x)$. If $k = \pi/a$ then the sine terms will all be zero, leaving just the cosine terms $\cos(n\pi)$, which can

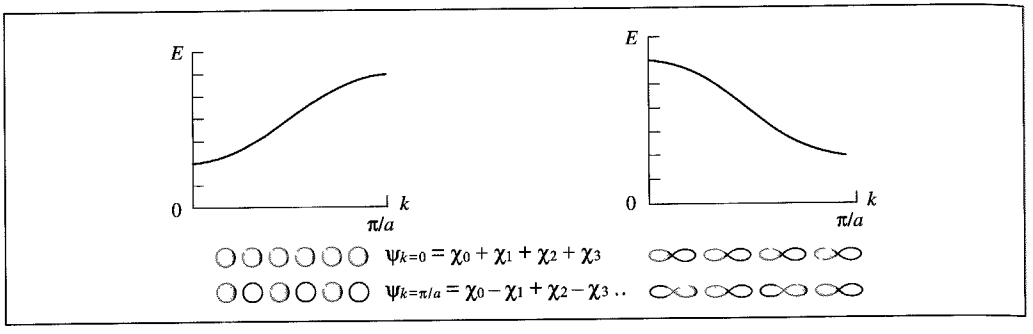

Fig 3 12 The variation in energy with k for a 1D lattice for a set of s orbitals (left) and for a set of p_x orbitals (right) Also shown are the corresponding arrangements of orbitals.

be expressed more generally as $(-1)^n$. Hence the wavefunction is:

$$\psi^{k=\pi/a} = \sum_n (-1)^n \chi_n = \chi_0 - \chi_1 + \chi_2 - \cdots \tag{3.84}$$

Equations (3.83) and (3.84) correspond to the lowest- and highest-energy wavefunctions for our simple system over this range of k. Wavefunctions for values of k between 0 and π/a have intermediate energies. The energy varies in a cosine-like manner with k between $k = 0$ and $k = \pi/a$ (Figure 3.12). Note that k can adopt negative values and that $E(-k)$ equals $E(k)$. Also worthy of note is that p orbitals show different behaviour to the s orbitals. For a set of p_x orbitals it is the $k = 0$ state that is of highest energy and $k = \pi/a$ is of lowest energy, due to their nodal behaviour

The graph of energy versus k is called the *band structure*; the *bandwidth* is the difference in energy between the lowest and highest levels in the band. For the one-dimensional lattice the bandwidth is determined by the lattice spacing; a smaller spacing a gives a greater bandwidth in much the same way that the difference between the bonding and antibonding orbitals in H_2 increases as the atoms get closer together. As we noted above there are as many values of k (and so as many energy levels) as there are atoms in the lattice and that each energy level can accommodate two electrons.

We now move on to consider a two-dimensional square lattice in the (x, y) plane, where the inter-lattice spacing is still a. The Bloch theorem is now written in the following more general form:

$$\psi^{\mathbf{k}}(\mathbf{r} + \mathbf{T}) = e^{i\mathbf{k}\,\mathbf{T}}\psi^{\mathbf{k}}(\mathbf{r}) \tag{3.85}$$

In Equation (3.85) **T** is a translation vector that maps each position into an equivalent position in a neighbouring cell, r is a general positional vector and **k** is the *wavevector* which characterises the wavefunction **k** has components k_x and k_y in two dimensions and is equivalent to the parameter k in the one-dimensional system. For the two-dimensional square lattice the Schrödinger equation can be expressed in terms of separate wavefunctions along the x- and y-directions. This results in various combinations of the atomic 1s orbitals, some of which are shown in Figure 3.13. These combinations have different energies. The lowest-energy solution corresponds to ($k_x = 0$, $k_y = 0$) and is a straightforward linear

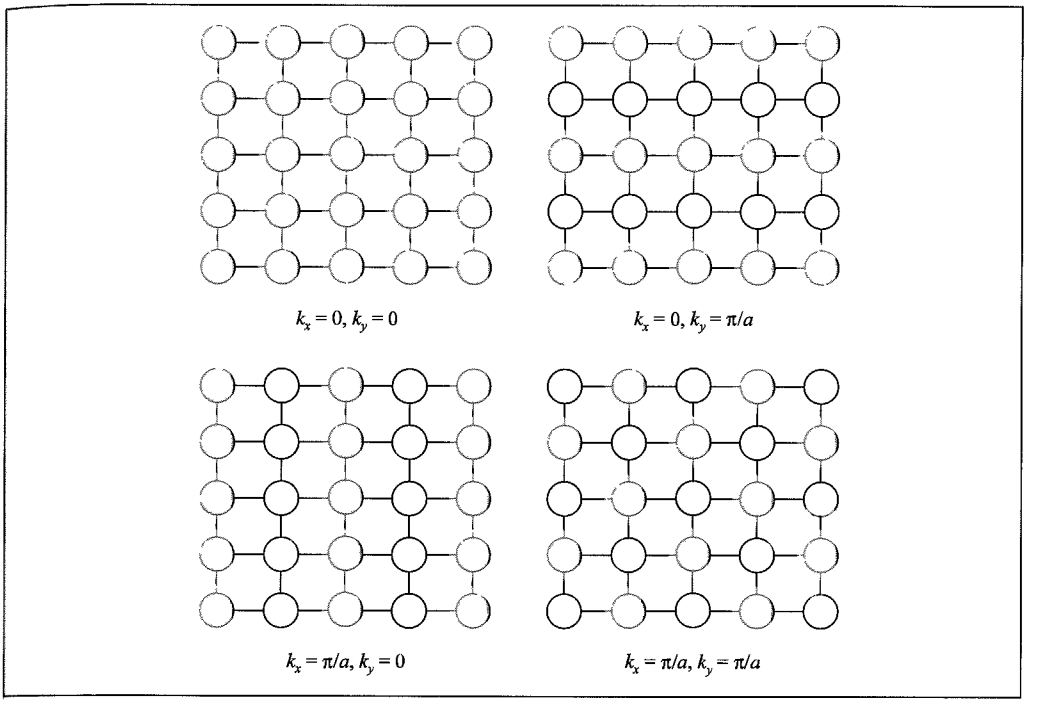

Fig 3.13. Some of the possible combinations of atomic 1s orbitals for a 2D square lattice corresponding to different values of k_x and k_y. A shaded circle indicates a positive coefficient; an open circle corresponds to a negative coefficient.

combination of the atomic orbitals The highest-energy solution corresponds to the situation where both k_x and k_y have values of π/a. The wavefunction for this high-energy solution shows a rapid variation in sign. Another important feature evident in Figure 3.13 is the wave-like nature of the various linear combinations, particularly if one imagines the lattice extending infinitely in all directions over the (x, y) plane.

The reciprocal space and the reciprocal lattice are directly related to the wavevector, $\mathbf{k}$; different values of $\mathbf{k}$ can be considered as points within the reciprocal space defined by $\mathbf{a}^{\$}$, $\mathbf{b}^{\$}$ and $\mathbf{c}^{\$}$. It turns out that, when we are calculating the wavefunction and energy levels for a solid, we need to restrict $\mathbf{k}$ to one cell in the reciprocal lattice (typically chosen to the cell containing $\mathbf{k} = 0$, or the first Brillouin zone), otherwise there is a danger of counting some states more than once. A very common way to represent the band structure for lattice structures is to plot how the energy changes as a function of $\mathbf{k}$ along certain lines of symmetry within the first Brillouin zone. For example, to return to our square lattice (for which the reciprocal lattice is also square) one could imagine taking a 'tour' starting at the origin ($\mathbf{k} = (0, 0)$), moving along the x axis to $\mathbf{k} = (\pi/a, 0)$ up the y axis to $\mathbf{k} = (\pi/a, \pi/a)$, and finally returning to the origin. As we undertake this tour the energy changes as shown in Figure 3.14. In this diagram we have labelled certain values of $\mathbf{k}$ which have particular symmetry with their conventional Roman or Greek capital letters, Γ, X and M [Bradley and Cracknell 1972].

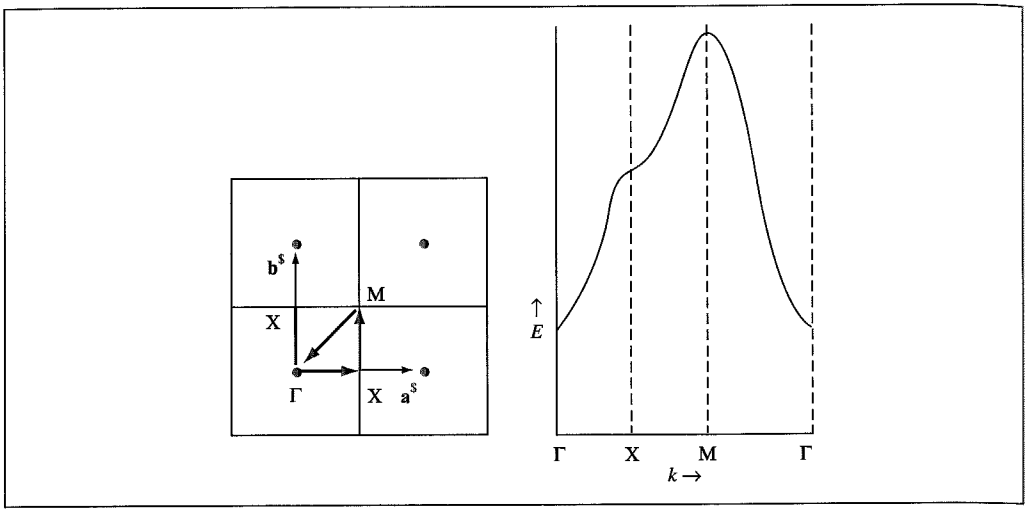

Fig. 3.14: Variation in energy for a 'tour' (Γ–X–M–Γ) of the reciprocal lattice for a 2D square lattice of hydrogen atoms. (Figure adapted in part from Hoffmann R 1988. Solids and Surfaces. A Chemist's View on Bonding in Extended Structures *New York, VCH Publishers.)*

3.8.3 The Periodic Hartree–Fock Approach to Studying the Solid State

In the *periodic Hartree–Fock* approach the elements of the Fock matrix are constructed from linear combinations of so-called Bloch functions:

$$\psi_i^{\mathbf{k}}(\mathbf{r}) = \sum_{\omega} a_{\omega i}(\mathbf{k})\varphi_{\omega}^{\mathbf{k}}(\mathbf{r}) \tag{3.86}$$

Each Bloch function is itself a linear combination of atomic orbitals:

$$\varphi_{\omega}^{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{T}} \chi_{\omega}^{\mathbf{T}}(\mathbf{r}) \exp(i\mathbf{k}\cdot\mathbf{T}) \tag{3.87}$$

$\chi_{\omega}^{\mathbf{T}}$ is the ωth atomic orbital in the crystal cell characterised by the lattice vector $\mathbf{T}$. As such, this method works in real space, which contrasts with the usual implementations of the alternative plane-wave methods that we will discuss below [Dovesi *et al.* 2000]. Each atomic orbital is expressed as a linear combination of (for example) Gaussian functions, as in molecular Hartree–Fock theory. The coefficients $a_{\omega i}(\mathbf{k})$ in Equation (3.86) are obtained by solving the following matrix equation for every value of $\mathbf{k}$ to self-consistency:

$$\mathbf{F}_{\mathbf{k}}\mathbf{A}_{\mathbf{k}} = \mathbf{S}_{\mathbf{k}}\mathbf{A}_{\mathbf{k}}\mathbf{E}_{\mathbf{k}} \tag{3.88}$$

$\mathbf{S}_{\mathbf{k}}$ is the overlap matrix for the Bloch functions for the wavevector $\mathbf{k}$, with $\mathbf{E}_{\mathbf{k}}$ being the energy matrix and $\mathbf{A}$ the matrix of coefficients. $\mathbf{F}_{\mathbf{k}}$ is the Fock matrix, which consists of a sum of one- and two-electron terms. The values of $\mathbf{k}$ are typically selected to sample from the first Brillouin zone according to a special scheme as described in Section 3.8.6. When these terms are expanded they involve infinite sums over the nuclei and electrons in the lattice. As is usual in a Hartree–Fock approach the one-electron terms involve the sum of a kinetic energy term and one due to the Coulomb interaction between the nuclei and the

electrons; the two-electron terms involve Coulomb and exchange two-electron integrals. Unfortunately, if these sums were to be evaluated individually and to completion then they would not converge to a consistent value, but would diverge. However, effective ways to determine these infinite sums have been proposed [Pisani and Dovesi 1980; Dovesi *et al.* 1983]. These involve a variety of procedures. The Coulomb interactions are divided into a series of terms corresponding to interacting and non-interacting charge distributions The latter can then be grouped together into 'shells' and the interaction calculated using multipole expansions (see Section 4.9.1). For the shorter-range exchange interaction it is possible to truncate the integral summation at an appropriate distance without loss of accuracy. The truncation distance can depend upon the three-dimensional structure of the material and so may vary from one calculation to the next.

Within the periodic Hartree–Fock approach it is possible to incorporate many of the variants that we have discussed, such as UHF or RHF. Density functional theory can also be used. This makes it possible to compare the results obtained from these variants. Whilst density functional theory is more widely used for solid-state applications, there are certain types of problem that are currently more amenable to the Hartree–Fock method. Of particular relevance here are systems containing unpaired electrons, two recent examples being the electronic and magnetic properties of nickel oxide and alkaline earth oxides doped with alkali metal ions (Li in CaO) [Dovesi *et al.* 2000].

3.8.4 The Nearly Free-electron Approximation

Whereas the tight-binding approximation works well for certain types of solid, for other systems it is often more useful to consider the valence electrons as free particles whose motion is modulated by the presence of the lattice. Our starting point here is the Schrödinger equation for a free particle in a one-dimensional, infinitely large box:

$$\left(\frac{d^2}{dx^2}\right)\psi = -\left(\frac{2mE}{\hbar^2}\right)\psi \tag{3 89}$$

The solutions to this equation are:

$$\psi = C\exp(ikx); \quad E = (\hbar^2 k^2)/2m \tag{3.90}$$

The energy for a free particle can be related to the momentum by $E = p^2/2m$ and so the wavefunction is related to the momentum p by:

$$\psi = C\exp(\pm ipx/\hbar) \tag{3.91}$$

The wavelength of this motion is h/p and the parameter k is equal to $2\pi p/h$. Thus k has units of 1/length (i.e. reciprocal length). The energy for a free particle varies in a quadratic fashion with k and in principle any value of the energy is possible.

In two dimensions we obtain the following wavefunction:

$$\psi_{x,y} = C_x \exp(ik_x x/\hbar) C_y \exp(ik_y y/\hbar) = C\exp(i\mathbf{k}\cdot\mathbf{r}/\hbar) \tag{3.92}$$

Note that in Equation (3.92) we have expressed the wavefunction in terms of a vector, $\mathbf{k}$ (which has components in the x and y directions of k_x and k_y) and the Cartesian vector $\mathbf{r}$.

The energy varies as a quadratic function of both k_x and k_y:

$$E_{x,y} = \frac{\hbar^2}{2m}(k_x^2 + k_y^2) \tag{3.93}$$

An analogous expression is obtained in three dimensions. We now need to consider periodic systems. As we have discussed, the wavefunction for a particle on a periodic lattice must satisfy Bloch's theorem, Equation (3.85). The wavevector $\mathbf{k}$ in Bloch's theorem plays the same role in the study of periodic systems as the vector $\mathbf{k}$ does for a free particle. One important difference is that whereas the wavevector is directly related to the momentum for a free particle (i.e. $\mathbf{k} = \mathbf{p}/\hbar$) this is not the case for the Bloch particle due to the presence of the external potential (i.e. the nuclei). However, it is very convenient to consider $\hbar\mathbf{k}$ as analogous to the momentum and it is often referred to as the *crystal momentum* for this reason. The possible values that $\mathbf{k}$ can adopt are given by:

$$\mathbf{k} = \left(\frac{m_\alpha}{N_\alpha}\mathbf{a}^\$, \frac{m_\beta}{N_\beta}\mathbf{b}^\$, \frac{m_\gamma}{N_\gamma}\mathbf{c}^\$\right) \tag{3.94}$$

m_α, m_β and m_γ are integers and $N_\alpha N_\beta N_\gamma = N$, the number of unit cells in the crystal. For a macroscopic system where N is very large (of the order of Avogadro's number) $\mathbf{k}$ thus varies continuously. As we have seen before, the wavevector $\mathbf{k}$ in the Bloch theorem (Equation (3.85)) can be considered as a point within the reciprocal lattice defined by $\mathbf{a}^\$$, $\mathbf{b}^\$$ and $\mathbf{c}^\$$. It can also be shown (see Appendix 3.1) that a wavefunction that satisfies Bloch's theorem can be written in the following form:

$$\psi^{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u^{\mathbf{k}}(\mathbf{r}) \tag{3.95}$$

Here, $u^{\mathbf{k}}(\mathbf{r})$ is a function that is periodic on the lattice. Recall from our earlier discussions on reciprocal lattice vectors that one way to construct such a periodic function is as a Fourier series expansion of plane wavefunctions $\exp(i\mathbf{G}\cdot\mathbf{r})$:

$$u^{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{G}}^{\mathbf{k}} \exp(i\mathbf{G}\cdot\mathbf{r}) \tag{3.96}$$

The sum runs over the reciprocal lattice vectors $\mathbf{G}$ we considered above A simple case is $\mathbf{G} = \mathbf{a}^\$$, for which $\exp(i\mathbf{G}\cdot\mathbf{r})$ corresponds to a wave travelling perpendicular to the real-space axes $\mathbf{b}$ and $\mathbf{c}$ and with a wavelength such that it fits exactly into the unit cell. If $\mathbf{G} = 2\mathbf{a}^\$$ then two wavelengths fit into the cell.

The external potential due to the nuclei is periodic in the lattice and it too can be written as a Fourier expansion of exponential functions of the reciprocal lattice:

$$U(\mathbf{r}) = \sum_{\mathbf{G}} U_{\mathbf{G}} \exp(i\mathbf{G}\cdot\mathbf{r}) \tag{3 97}$$

$U_{\mathbf{G}}$ is the Fourier coefficient. When this form of the potential is incorporated into the Schrödinger equation the following equation can be derived [Ashcroft and Mermin 1976]:

$$\left(\frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}|^2 - E\right)c_{\mathbf{G}}^{\mathbf{k}} + \sum_{\mathbf{G}'} U_{\mathbf{G}'+\mathbf{G}}\, c_{\mathbf{G}'}^{\mathbf{k}} = 0 \tag{3.98}$$

We can recover the free-particle result (i.e. zero potential) from Equation (3.98) by setting all of the Fourier coefficients $U_{\mathbf{G}}$ to zero, in which case the equation reduces to:

$$\left(\frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}|^2-E\right)c_{\mathbf{G}}^{\mathbf{k}}=0 \tag{3.99}$$

The solution of this equation requires that $E=\hbar^2|\mathbf{k}+\mathbf{G}|^2/2m$ with the wavefunctions being of the form $\psi(\mathbf{r})\propto\exp[i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}]$. Although cast in a slightly different form, this is equivalent to our earlier expression for the wavefunction of a free particle, Equation (3.92).

The summations in Equations (3.98) are over all reciprocal lattice vectors **G**. As can be seen, for a given value of **k** there are as many forms of this equation as there are reciprocal lattice vectors in the system. Each of these equations for the different values of **G** gives rise to a solution which is labelled with the band index n. Thus there are as many values of n as there are reciprocal lattice vectors **G**. Just as there are n solutions to this Schrödinger equation for a given value of **k**, so it is also possible to consider how the energy varies with **k** for a given value of n. To understand the entire band structure of a solid requires one to consider the variation of both **k** and n. As we indicated above, when calculating the band structure it is usual to restrict **k** to just the first Brillouin zone to avoid duplicate counting of states.

Let us now examine how these results can be applied to some simple one- and two-dimensional periodic systems. Initially we will consider the situation where there is no external potential and then discuss what happens when we introduce one. The first case is the one-dimensional lattice, which has reciprocal lattice vectors at $\pm2\pi/a$, $\pm4\pi/a$, etc. In order to derive the energy diagram we need to consider, for each reciprocal lattice vector **G**, how the energy varies as we change **k** over the first Brillouin zone (which in this case corresponds to varying **k** from $-\pi/a$ to $+\pi/a$). The first reciprocal lattice vector is $\mathbf{G}=0$, for which the energy simply varies quadratically with **k**, from zero at $\mathbf{k}=0$ to $\hbar^2(\pi/a)^2/2m$ at $\mathbf{k}=\pm2\pi/a$. We next need to consider the two reciprocal vectors $\mathbf{G}=\pm2\pi/a$. At the point $\mathbf{k}=0$ the energy due to both of these reciprocal lattice vectors is $\hbar^2(2\pi/a)^2/2m$. As **k** increases from 0 to $+\pi/a$ the value of $|\mathbf{k}+\mathbf{G}|^2$ increases for the reciprocal lattice vector $\mathbf{G}=2\pi/a$ but it decreases for the reciprocal lattice vector $\mathbf{G}=-2\pi/a$. Conversely, as **k** varies from 0 to $-\pi/a$ the energy increases for the reciprocal lattice vector $\mathbf{G}=-2\pi/a$ and decreases for $\mathbf{G}=2\pi/a$. These variations in energy are shown in Figure 3.15. Two types of energy diagram are shown in this figure; one is the 'reduced-zone' scheme because the entire dependency of the energy on the wavevector is contained within the first Brillouin zone. The alternative representation is called an extended-zone scheme in which the energy levels are 'folded out' for values of **k** beyond the first Brillouin zone.

We next need to introduce the weak potential, which acts to modulate the wavefunctions and the associated energy levels. The effects of the potential are found to be most acute where there is degeneracy of the energy levels. This arises even in the one-dimensional situation, where we have degenerate energy levels due to different reciprocal lattice vectors at $\mathbf{k}=0$ and $\mathbf{k}=\pi/a$. The effect of the potential is to perturb these energy levels in such a way that lifts the degeneracy to create an energy gap. In the one-dimensional case the effect of the potential is to 'flatten' the energy levels in the region close to the edge of the

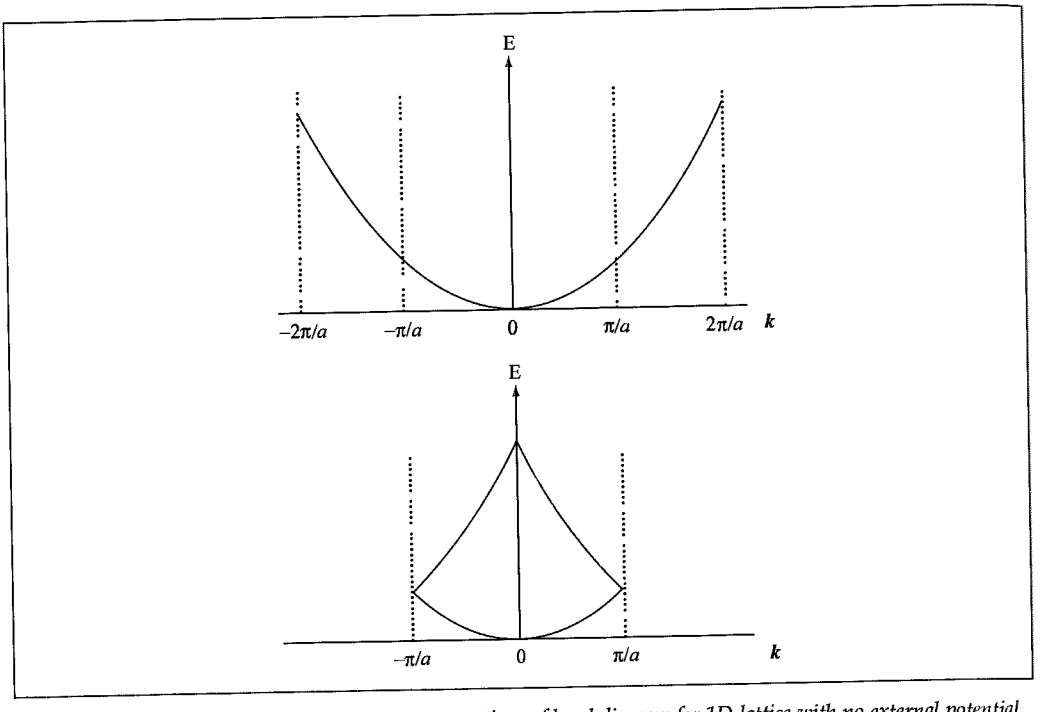

Fig 3.15. Extended-zone and reduced-zone representations of band diagram for 1D lattice with no external potential.

Brillouin zone as shown in Figure 3.16. One way to explain the appearance of the energy gap at the edges of the Brillouin zone is to recognise that the states of a free electron are waves with a specific wavelength ($2\pi/k$ in the simple one-dimensional system). When the wavelength becomes comparable to the lattice spacing the lattice diffracts the wave and at the boundary of the Brillouin zone ($k = \pm\pi/a$) a standing wave is created. Two different standing waves are possible in a one-dimensional system, as shown in Figure 3.17. For one of the standing waves (A in Figure 3.17) the peak electron density occurs in the vicinity of the lattice points (the positive nuclei). This standing wave thus has a more favourable (i.e. lower) energy than the equivalent free travelling wave. By contrast, the peak electron density of the other standing wave (B in Figure 3.17) occurs between the nuclei and so its energy is higher. Further gaps arise at $k = \pm 2\pi/a$, and so on.

A somewhat more complex case is that of the 2D hexagonal lattice. As for the one-dimensional system we initially consider a free particle, restricting ourselves to wavevectors within the first Brillouin zone with higher-energy states being due to reciprocal lattice vectors beyond in the second, third, etc. Brillouin zones. We will consider how the energy varies as we undertake a 'tour' of the first Brillouin zone in reciprocal space starting at the origin ($\mathbf{k} = (0,0)$), then moving to one of the vertices of the hexagon (the point ($\mathbf{k} = \cos\pi/6, \sin\pi/6$)), along to the mid-point of one of the edges ($\mathbf{k} = (0, \sin\pi/6)$), and finally back to the origin (Figure 3.18). The origin, the vertex and the mid-point are all points of symmetry and are identified by the symbols Γ, K and M, respectively. For a

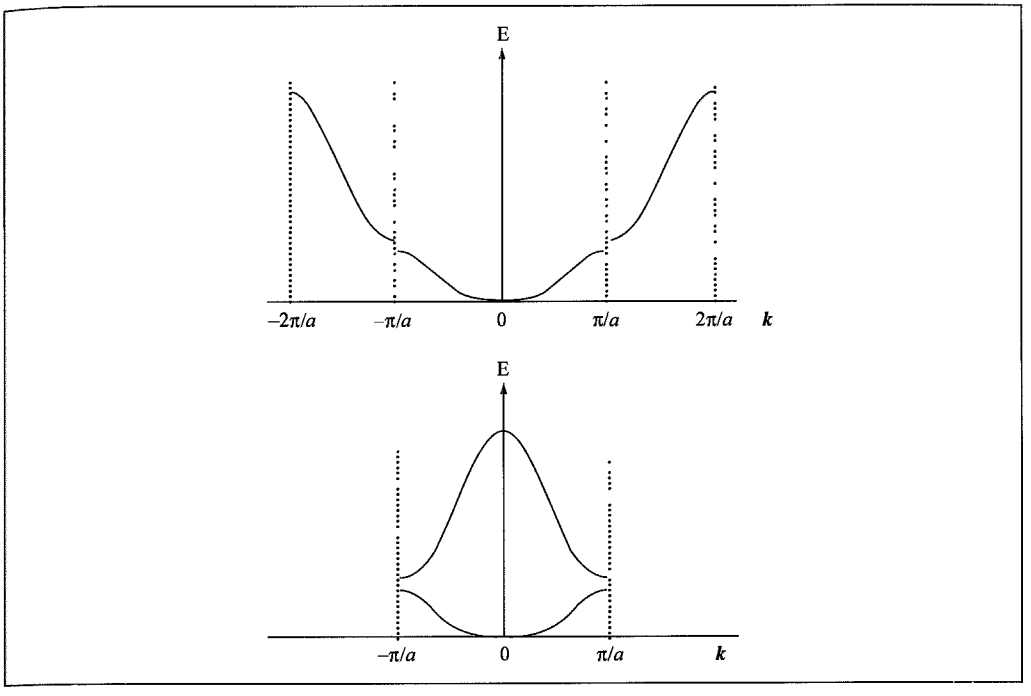

Fig 3 16: The effect of introducing a weak potential into the 1D lattice is to lift the degeneracy of the energy levels near to the edge of the Brillouin zone (shown in both extended-zone and reduced-zone representation)

given value of $\mathbf{k}$ we compute the value of $|\mathbf{k}+\mathbf{G}|^2$ and thus the energy for the relevant reciprocal lattice vectors.

The simplest case is that corresponding to $\mathbf{G} = 0$. We still obtain a quadratic variation of energy with $|\mathbf{k}|$ wherever we move within the first Brillouin zone. The variation in energy for the three 'legs' of this tour can be represented in an energy band diagram as shown in Figure 3.18. As there are six nearest-neighbour cells in this system, there are six energy levels to monitor at the next stage. The distance from the origin to each of these six reciprocal lattice points is $2\cos\pi/6$. At $\mathbf{k} = 0$ we therefore find that all six energy levels are degenerate

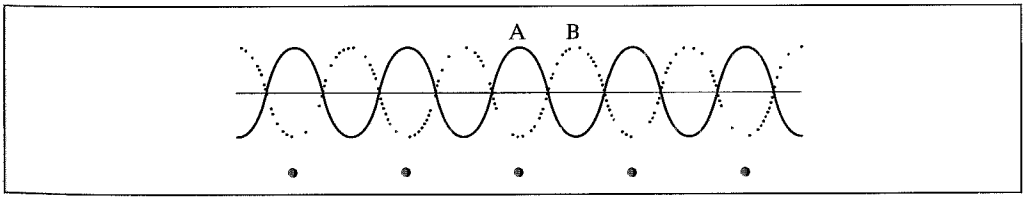

Fig 3 17 The two possible sets of standing waves at the Brillouin zone boundary Standing wave A concentrates electron density at the nuclei, whereas wave B concentrates electron density between the nuclei. Wave A thus has a lower energy than wave B

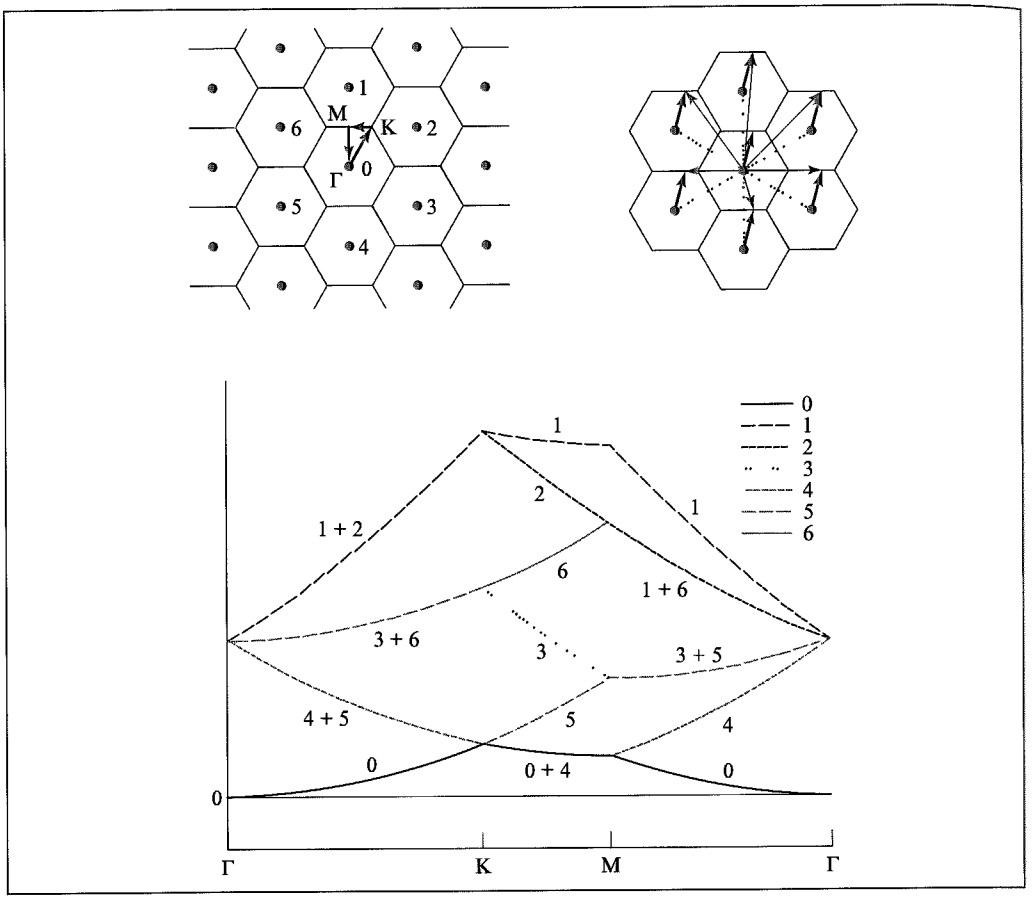

Fig 3.18: Energy band diagram (bottom) for a free-particle 'tour' (Γ–K–M–Γ) of the reciprocal lattice for a 2D hexagonal structure (top left) A total of seven bands are shown, due to the central reciprocal lattice vector $\boldsymbol{G} = \boldsymbol{0}$ *and the reciprocal lattice vectors from the six neighbouring cells The energy varies as* $|\boldsymbol{k} + \boldsymbol{G}|^2$*, where the vector* $\boldsymbol{k} + \boldsymbol{G}$ *is computed as shown in the top right of the figure (*$\boldsymbol{k}$*, bold arrow;* $\boldsymbol{k} + \boldsymbol{G}$*: thin arrow).*

and have a value of $3\hbar^2/2m$ ($(2\cos\pi/6)^2 \equiv 3$). Moving towards the point $(\cos\pi/6, \sin\pi/6)$ we find that the six vectors separate into three pairs of degenerate levels. These six reciprocal lattice points are labelled 1–6 in Figure 3.18, together with the corresponding energy levels. As the tour continues, the different energy bands show two-, three- and six-fold degeneracy, depending upon the value of k. Another key feature is that along some legs of the tour certain pairs of bands are degenerate, though this degeneracy will often be lifted when a different leg is traversed. For example, the pairs 1–2, 3–6 and 4–5 are degenerate from Γ to K. Between K and M the pair 0–4 are degenerate; and on the final leg there is degeneracy between the pairs 2–6 and 3–4. When the periodic potential is introduced some, but not necessarily all, of this degeneracy will be lifted, giving rise to band gaps. The way in which this can occur is shown schematically in Figure 3.19.

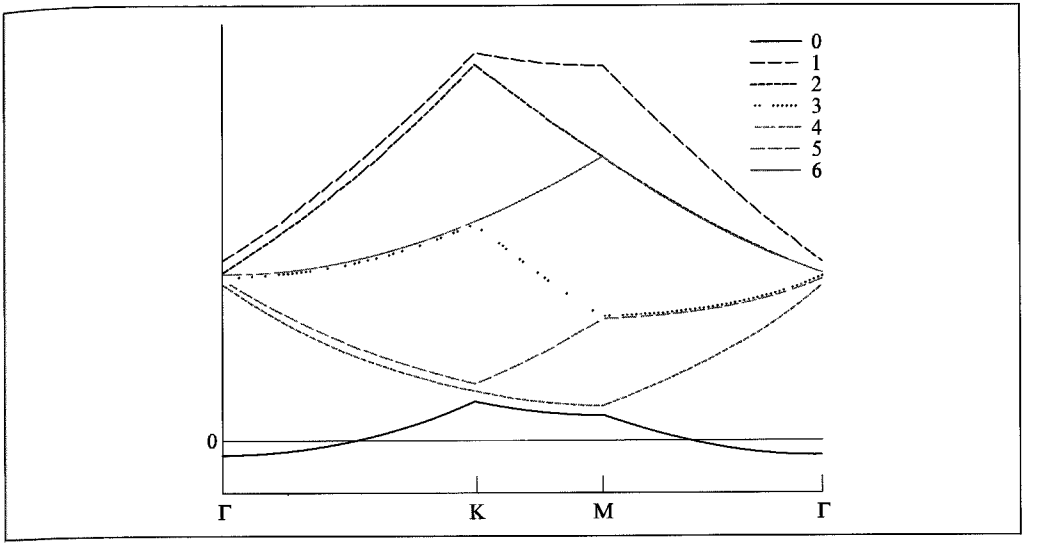

Fig. 3 19. The effect of a weak external potential is to lift degeneracy and create band gaps as illustrated for a 2D hexagonal lattice (compare with Figure 3.18).

3.8.5 The Fermi Surface and Density of States

To determine the ground state of a periodic system it is necessary to determine its band structure, by varying **k** over the first Brillouin zone and computing at each value of **k** the different energy bands resulting from the reciprocal lattice vectors. The number of energy levels in a band (i e. the number of values permitted to **k**) is equal to the number of primitive cells in the crystal, just as was the case for the orbital model in the tight-binding approximation. For each energy level corresponding to a particular value of **k** the Pauli principle permits two electrons of opposite spin to be assigned. This process is repeated for the different bands until all the electrons have been allocated. The energy level of the highest occupied state is called the *Fermi energy* (for a metal; for an insulator, the Fermi energy is in the middle of a band gap). When all the electrons have been assigned then one of two different situations may result. In the first case all the occupied bands are completely filled. As we saw earlier, this gives rise to a band gap between the top of the highest occupied level and the bottom of the lowest empty level. The number of energy levels in each band is equal to the number of primitive cells in the crystal, so a band gap can only arise if there is an even number of electrons per primitive cell. The tight-binding approximation discussed in Section 3.8 2 may be an appropriate model to apply in this case. The second situation arises when one or more bands are partially filled. For each of these partially filled bands one can consider there to be a surface in the **k** space that separates the occupied and the unoccupied levels, as defined by the Fermi energy. This set of surfaces is known as the *Fermi surface* and it defines a border between the occupied and unoccupied states. In many cases the Fermi surface is contained within a single band; if not, then the parts of the Fermi surface due to partially filled individual bands are known as the *branches* of the Fermi surface The

Fermi surface will show the same underlying periodicity as the reciprocal lattice. A particularly attractive feature of the Fermi surface is that it can be measured experimentally, so providing a link between theory and experiment.

The *density of levels* is another useful way to describe the electronic structure of a solid. The density of levels indicates how many energy levels there are for a particular energy. It can thus be defined as the number of levels between E and $E + dE$. This is often normalised by volume, leading to the density of levels per unit volume $g(E)$, which is given by:

$$g(E) = \sum_n g_n(E) \tag{3.100}$$

The sum is over the bands n, with $g_n(E)$ being the density of levels in the band n:

$$g_n(E) = \frac{1}{4\pi^3} \int \delta(E - E_n(\mathbf{k}))\, d\mathbf{k} \tag{3 101}$$

The delta function $\delta(E - E_n(\mathbf{k}))$ has a value of 1 if $E_n(\mathbf{k})$ is in the range E to $E + dE$ and 0 otherwise. The density of states $D(E)$ is closely related to the density of levels; in the simple case where we have two electrons in each level then the density of states is just twice the density of levels. The integral of the density of states up to the Fermi level is equal to the number of electrons and the integral of the density of states multiplied by the energy is the total electronic energy:

$$N = \int D(E)\, dE \tag{3.102}$$

$$E_{\text{tot}} = \int D(E) E\, dE \tag{3 103}$$

The density of states can be usefully visualised by plotting the energy versus $D(E)$. For the simple one-dimensional situation where the energy varies in a cosine-like manner with $\mathbf{k}$ and the levels are equally spaced, the density of states is greatest at the top and bottom of the band (Figure 3.20). The density of states is thus inversely proportional to

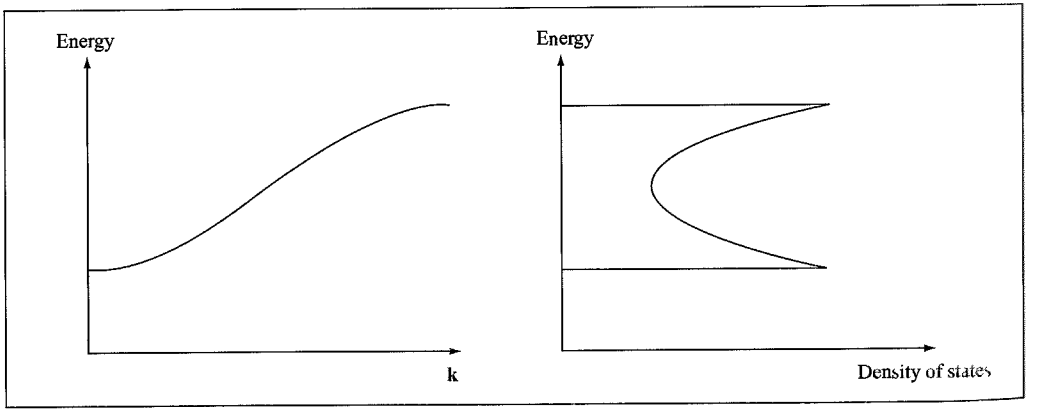

Fig 3.20. Variation of the density of states, D(E), for the simple 1D lattice, shown with the corresponding energy diagram

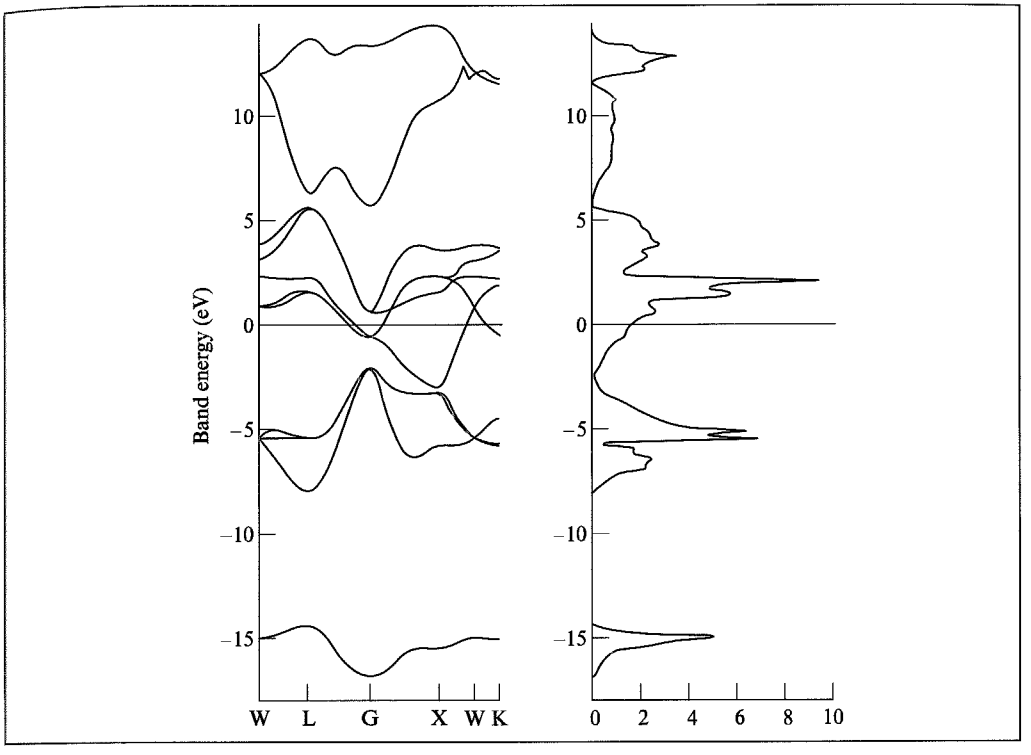

Fig. 3.21. Band structure and density of states for TiN.

the slope of the energy versus **k** curve; the flatter the band the greater the density of states at that energy.

The density of states is somewhat like an orbital energy diagram, but unlike the latter does not contain well-defined individual energy levels. Nevertheless, in some situations it is possible to determine from which atomic orbitals a particular energy band is largely derived. Of course, most real systems have rather more complex electronic structures than the simple cases we have used to discuss the background, as illustrated in Figure 3.21, which shows the band structure and density of states diagram for TiN.

3.8.6 Density Functional Methods for Studying the Solid State: Plane Waves and Pseudopotentials

Plane waves are often considered the most obvious basis set to use for calculations on periodic systems, not least because this representation is equivalent to a Fourier series, which itself is the natural language of periodic functions. Each orbital wavefunction is expressed as a linear combination of plane waves which differ by reciprocal lattice vectors:

$$\psi_i^{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} a_{i,\mathbf{k}+\mathbf{G}} \exp(i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}) \tag{3.104}$$

The Kohn–Sham equations of the density functional theory then take on the following form:

$$\sum_{G'}\left\{\frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}|\delta_{GG'}+V_{\text{ion}}(\mathbf{G}-\mathbf{G}')+V_{\text{elec}}(\mathbf{G}-\mathbf{G}')+V_{\text{XC}}(\mathbf{G}-\mathbf{G}')\right\}a_{i,\mathbf{k}+\mathbf{G}'}=\varepsilon_i a_{i,\mathbf{k}+\mathbf{G}'} \tag{3.105}$$

V_{ion}, V_{elec} and V_{XC} represent the electron–nuclei, electron–electron and exchange-correlation functionals, respectively. The delta function $\delta_{GG'}$ is zero unless $\mathbf{G}=\mathbf{G}'$, in which case it has a value of 1 There are two potential problems with the practical use of this equation for a 'macroscopic' lattice. First, the summation over $\mathbf{G}'$ (a Fourier series) is in theory over an infinite number of reciprocal lattice vectors. In addition, for a macroscropic lattice there are effectively an infinite number of $\mathbf{k}$ points within the first Brillouin zone. Fortunately, there are practical solutions to both of these problems.

We are usually interested in the valence electrons of an atom, as these are largely responsible for the chemical bonding and most physical properties. The core electrons are little affected by the atomic environment. It is therefore common only to consider explicitly the valence electrons in the calculation and to subsume the core electrons into the nuclear core. One potential drawback to the representation of valence electron wavefunctions with a plane-wave basis set is that near to the atomic nuclei the wavefunctions of the valence electrons show rapid oscillations. This is because their wavefunctions must be orthogonal to those of the core electrons. These oscillations give rise to a large kinetic energy, and a very large number of plane waves would be required to properly model this behaviour. This corresponds to taking many terms in the plane-wave expansion of the orbital, Equation (3.104). This problem is compounded by the fact that the solid systems of interest often contain elements much later in the periodic table than are usually encountered in molecular Hartree–Fock calculations. Heavy elements have many more core electrons and so an even more pronounced oscillatory behaviour. However, in this inner region the kinetic energy is largely cancelled by the high electrostatic potential energy of interaction with the nucleus A popular way to deal with these problems is to replace the 'true' potential in these core regions with a much weaker one called a *pseudopotential*. This represents the way in which the valence electrons interact with the combined nucleus plus core electrons [Heine 1970]. A pseudopotential is a potential function that gives wavefunctions with the same shape as the true wavefunction outside the core region but with fewer nodes inside the core region, as illustrated in Figure 3.22. This has the effect of reducing the number of terms required for the plane wave expansion of the wavefunction, which in turn drastically reduces the scale of the computational problem.

Pseudopotentials are usually derived from all-electron atomic calculations. The valence electron pseudopotential is then required to reproduce the behaviour and properties of the valence electrons in the full calculation. For example, the energy levels with the pseudopotential should be the same as for the all-electron calculation. In addition, the pseudopotential will often depend upon the orbital angular momentum of the wavefunction (i.e. for s, p, d, etc. orbitals) and it will be desired that the total valence electron density within the core radius equals that in the all-electron situation. Such pseudopotentials are

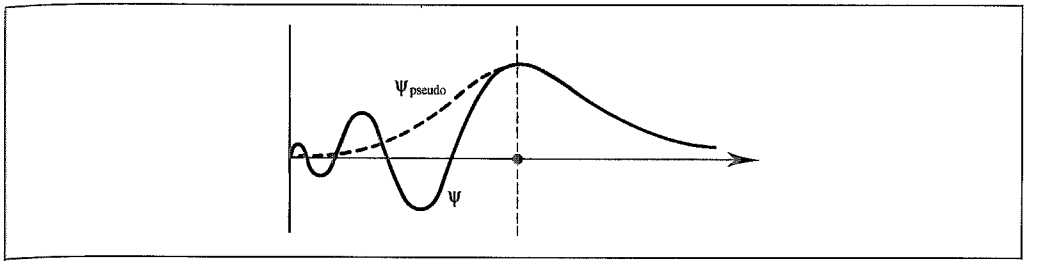

Fig. 3 22 Schematic representation of a pseudopotential (Figure adapted from Payne M C, M P Teter, D C Allan, R A Arias and D J Joannopoulos 1992 Iterative Minimisation Techniques for Ab initio *Total-Energy Calculations· Molecular Dynamics and Conjugate Gradients.* Reviews of Modern Physics *64 1045–1097.)*

referred to as 'non-local norm-conserving'. An additional advantage of the use of pseudopotentials for the heavy elements is that they enable some relativistic effects to be included in the model. A number of functional forms are possible for the pseudopotentials; it is usual to assume a specific functional form and then to vary the parameters. The various pseudopotentials differ in the number of plane waves that are required for their representation and in the degree to which they can be transferred between different atomic environments. So-called 'soft' pseudopotentials require fewer plane waves and are therefore computationally more attractive, though there is to some extent a trade-off between softness and transferability. Subsequently developed were the 'ultrasoft' or 'supersoft' pseudopotentials, which require even fewer plane waves.

In practice, therefore, a pseudopotential is invariably employed and only plane waves with a kinetic energy ($= (\hbar^2/2m)|\mathbf{k} + \mathbf{G}|^2$) less than some cutoff are included in the calculation The cutoff used depends on the nature of the system under investigation. For example, in the first-row elements the 2p valence orbitals approach closer to the nucleus than the comparable 3p orbitals in the second-row elements (the latter are repelled by the lower 2p states). Thus elements such as silicon or sulphur usually have softer pseudopotentials than their first-row equivalents carbon and oxygen. Everything else being equal, a higher cutoff is consequently required for the latter and hence more plane waves in the expansion (i.e. more reciprocal lattice vectors, **G**). Note that in the plane wave expansion the basis functions are not associated with particular atoms but are defined over the whole cell (this also removes the problem of basis-set superposition errors as an additional benefit). The coefficients $a_{i,\mathbf{k}+\mathbf{G}}$ are obtained by following the usual density functional scheme: an initial guess is made of the electron density variation $\rho(\mathrm{r})$, the Kohn–Sham and overlap matrices are constructed, diagonalisation gives the eigenfunctions and eigenvectors (and thus the coefficients a) from which the Kohn–Sham orbitals can be constructed and hence the density for the next iteration.

The second important practical consideration when calculating the band structure of a material is that, in principle, the calculation needs to be performed for all **k** vectors in the Brillouin zone This would seem to suggest that for a macroscopic solid an infinite number of vectors **k** would be needed to generate the band structure. However, in practice a discrete sampling over the Brillouin zone is used. This is possible because the wavefunctions at points

that are close together in **k** space will be almost identical and can be represented by a single representative point. Each of these discrete values is multiplied by a weight factor related to the volume of reciprocal space it represents. Obviously, the denser the set of **k** vectors the smaller will be the error in the calculation Various schemes have been suggested for selecting suitable sets of **k** vectors which can give very accurate approximations to properties such as the charge density; the method of Monkhorst and Pack is particularly popular [Monkhorst and Pack 1976]. The selection of **k** vectors is also influenced by the size and shape of the system; indeed, if the unit cell is large then it may only be necessary to consider just one vector. Typically, between ten and 100 vectors are sufficient to understand the structural and electronic properties of a solid, though for certain types of problem such as calculating the optical properties of a metal many more **k** vectors may be required (several thousands). Ideally, one should ensure that the calculation converges both in terms of the number of wave-vectors **k** considered and in terms of the number of reciprocal lattice vectors **G**. An additional consideration is that the symmetry of the Brillouin zone itself may mean that it is not necessary for **k** to vary over the entire zone but that only a smaller section need be considered For example, in our two-dimensional hexagonal close-packed case we would only have to consider the small right-angled triangle over which we undertook our 'tour'. This has an area one-twelfth that of the entire zone. This is an example of the use of the point symmetry of the Brillouin zone rather than the translational symmetry of the lattice. The small section containing the explicit **k** vectors required for the calculation is sometimes referred to as the *irreducible* part of the Brillouin zone

3.8.7 Application of Solid-state Quantum Mechanics to the Group 14 Elements

The combination of density functional methods with pseudopotentials has been used extensively to study a wide variety of materials. Three systems that have been the subject of much interest are the group 14 elements carbon, silicon and germanium, reflecting their natural abundance, commercial importance (especially for silicon) and the large amount of experimental data available. Of particular interest is the problem of predicting the lowest-energy structure at a given volume [Cohen 1986; Mujica and Needs 1993, Needs and Mujica 1995]. In effect, this corresponds to predicting the most stable structure at a particular pressure. These elements all exist in the familiar diamond structure at normal pressures and temperatures but alternative structures can be formed by the application of pressure, at least for silicon and germanium. There has also been much speculation as to whether diamond itself could be transformed should a high enough pressure be applied. This last problem does have some practical interest as it would provide a theoretical upper limit to the pressures that could be achieved with ultra high-pressure diamond anvil cells.

There are many alternatives to the diamond structure, including body-centred cubic, face-centred cubic, hexagonal close-packed, simple hexagonal, simple cubic, β-tin, double-hexagonal close-packed and two complex tetrahedral structures. a body-centred cubic structure with eight atoms per unit cell and a simple tetragonal structure with twelve atoms per unit cell, not forgetting of course the many fullerene forms. Not all studies

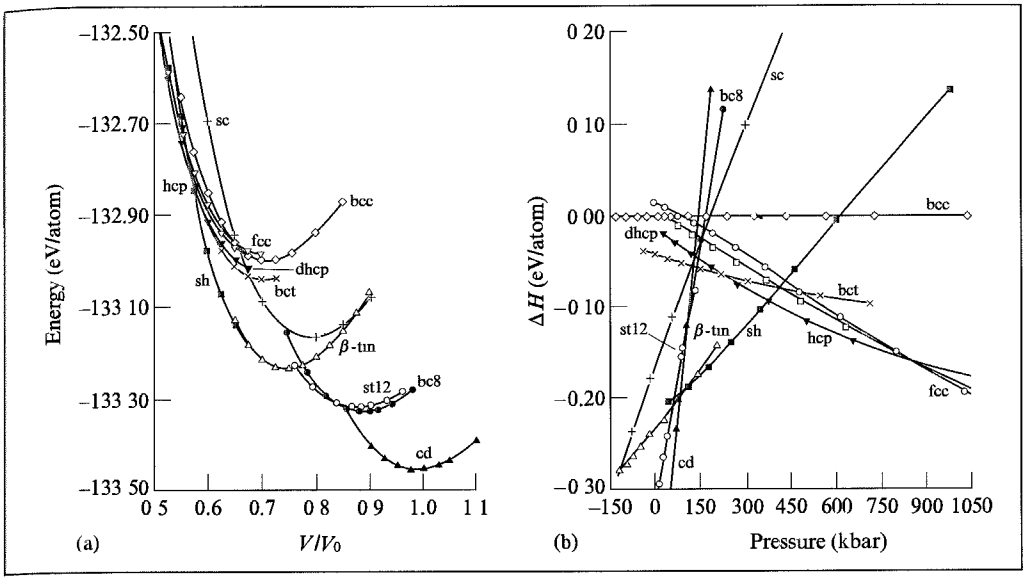

Fig. 3 23. (a) Graph of energy vs volume (scale normalised to the diamond structure) for eleven phases of silicon (b) Enthalpy–pressure plot for the same eleven phases relative to the body-centred cubic phase (Figures redrawn from Needs R J and A Mujica 1995. First-principles pseudopotential study of the structural phases of silicon. Physical Review ***B51***·*9652–9660.)*

consider every one of these phases but by quoting the list in full we can appreciate the range of possibilities. The energy differences between many of these phases are often small and so it is particularly important to achieve an effective sampling of points in **k** space (recent studies suggest several thousands of such points are needed). The plane-wave cutoff can also have an effect on the results. The calculations involve minimising each structure at a number of different volumes and then fitting a polynomial to the data points. The results are usually displayed as a graph of the total energy versus the volume, as shown in Figure 3 23. Another way to display this type of data is an enthalpy–pressure plot, from which the most stable phase at any pressure is easily identified as that with the lowest enthalpy. Various bulk structural properties can also be calculated for comparison with experiment.

As we alluded, of the forms mentioned above only the diamond structure has been observed experimentally for carbon. For both silicon and germanium there is a transition to the β-tin phase around 100–130 kbar. Silicon further transforms into other structures such as the simple hexagonal with a relatively modest further increase in pressure, whereas for germanium this transition requires much more pressure. Why should this be, given that they are all in the same group? The electronic structure calculations provide some significant insights into this problem. Thus silicon has a strongly repulsive p-orbital pseudopotential due to the inner (2p) electrons, which carbon does not. This repulsion contributes to the formation of a single peak in the electron density along each Si–Si bond, whereas for carbon there are two peaks, each being near the position for the atomic p orbitals (Figure 3.24). The differences

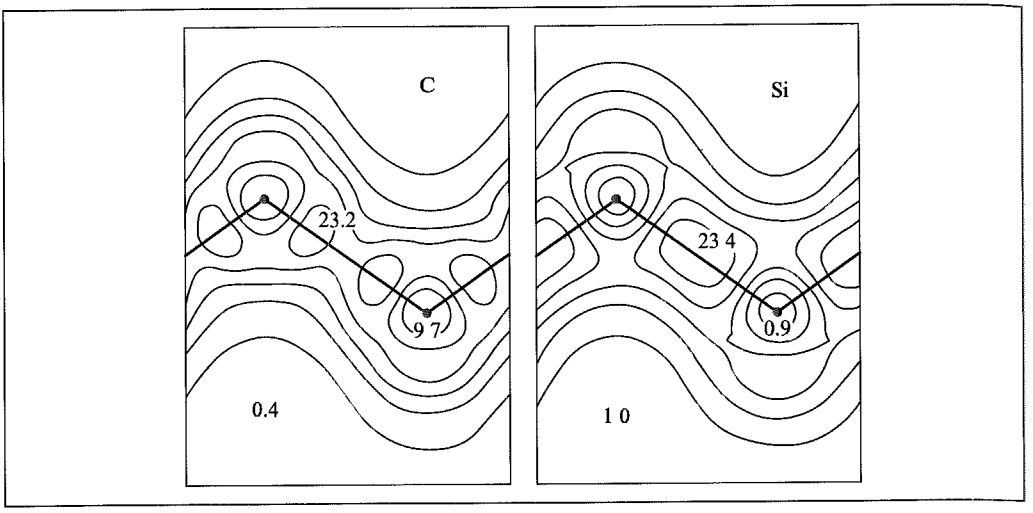

Fig 3 24: Valence electron density for the diamond structures of carbon and silicon (Figure redrawn from Cohen M L 1986 Predicting New Solids and Superconductors. Science **234** *549–553)*

between silicon and germanium are ascribed to the d electron states; silicon does not have core d electrons, whereas germanium does. Certain transitions (e.g. carbon $\rightarrow$ β-tin) do not depend upon the d character of the electronic configuration in contrast to subsequent transitions.

3.9 The Future Role of Quantum Mechanics: Theory and Experiment Working Together

Of all the methods that we will discuss in this book, quantum mechanics is probably the most widely used and the most extensively developed. The importance of the subject can be gauged in many ways, from citation counts to the number of Nobel prizes awarded. The systems studied using quantum mechanics range from the simplest molecular species (e.g. H_2^+, HD^+, H_3^+) to some very large and complex molecules (e.g. DNA, proteins and complex solid-state materials). Some of the most productive situations occur when experiment and theory are used in combination to tackle a problem. The methylene molecule, CH_2, is of particular historical interest. Despite its small size, this molecule and the controversy surrounding it played an important role in establishing the role of computational quantum mechanical methods in modern-day research and the relationship between theory and experiment [Schaeffer 1986]. The early debate concentrated on the ground state of the molecule and whether its geometry was linear or bent. Early *ab initio* calculations by Foster and Boys [Foster and Boys 1960] suggested an H-C-H angle of 129° but this was refuted by spectroscopic data from Herzberg's laboratory, which were interpreted to indicate a linear geometry Unfortunately for Foster and Boys, empirical calculations favoured by their head of department, Longuet-Higgins, also gave a linear geometry. Events came to a head when Bender and Schaeffer calculated a geometry of 135.1° and concluded that

the energy barrier between the linear and bent geometries was so large that no further improvement in the theoretical model could remove it. Soon thereafter several other experiments were undertaken, showing a bent structure. Moreover, when Herzberg re-examined his original data it was found to be consistent with a bent model. As we shall see in the remaining chapters there are many kinds of problem that can be tackled using computational chemistry methods. By no means do they always work, but there is often a synergistic relationship between experiment and theory, which means that the two combined can be much more productive than either in isolation.

Appendix 3.1 Alternative Expression for a Wavefunction Satisfying Bloch's Function

We have Equation (3.81):

$$\psi^k(x+a) = e^{ika}\psi^k(x) \tag{3.106}$$

We write $\psi(x)$ as the product of the exponential and a function $u_k(x)$:

$$u_k(x) = \psi_k(x)/\exp(ikx) \tag{3 107}$$

If we perform the same manipulation for $\psi(x+a)$ we get:

$$u_k(x+a) = \frac{\psi_k(x+a)}{e^{ik(x+a)}} = \frac{\psi_k(x)\,e^{ika}}{e^{ikx}\,e^{ika}} = \frac{\psi_k(x)}{e^{ikx}} = u_k(x) \tag{3.108}$$

Thus $u_k(x)$ is a periodic function which can be used to formulate acceptable wavefunctions:

$$\psi_k(x) = e^{ikx}u_k(x) \tag{3.109}$$

Further Reading

Ashcroft N W and N D Mermin 1976. *Solid State Physics* New York, Holt, Rinehart and Winston.

Atkins P W 1991 *Quanta. A Handbook of Concepts* Oxford, Oxford University Press.

Atkins P W and R S Friedman 1996. *Molecular Quantum Mechanics*, 3rd edition. Oxford, Oxford University Press.

Catlow C R A 1997. Computer Modelling as a Technique in Materials Chemistry In Catlow C R A and A K Cheetham (Editors). *New Trends in Materials Chemistry*, NATO ASI Series C 498, Dordrecht, Kluwer.

Catlow C R A 1998. Solids Computer Modelling. In Schleyer, P v R, N L Allinger, T Clark, J Gasteiger, P A Kollman, H F Schaeffer III and P R Schreiner (Editors) *The Encyclopedia of Computational Chemistry*, Chichester, John Wiley & Sons

Gillan M J 1991 Calculating the Properties of Materials from Scratch In Meyer M and V Pontikis (Editors) *Computer Simulation*, NATO ASI Series E 205 (Computer Simulations in Materials Science) pp 257–281.

Hehre W J, L Radom, P v R Schleyer and J A Pople 1986. *Ab initio Molecular Orbital Theory* New York, John Wiley & Sons

Hoffmann R 1988. *Solids and Surfaces: A Chemist's View on Bonding in Extended Structures.* New York, VCH Publishers.

Kohn W, A D Becke and R G Parr 1996. Density Functional Theory of Electronic Structure. *Journal of Chemical Physics* **100**:12974–12980.

Kohn W and P Vashita 1983. General Density Functional Theory. In Lundquist S and N H March (Editors). *Theory of Inhomogeneous Electron Gas,* New York, Plenum, pp. 79–148.

Kutzelnigg W and P von Herigonte 2000. Electron Correlation at the Dawn of the 21st Century *Advances in Quantum Chemistry* **36**:185–229.

Pisani C, R Dovesi and C Roetti 1988 Hartree-Fock *Ab initio* Treatment of Crystalline Systems. *Lecture Notes in Chemistry* Vol. 48. Berlin, Springer-Verlag

Pisani C, R Dovesi, C Roetti, M Cansa, R Orlando, S Casass and V R Saunders 2000. CRYSTAL and EMBED, Two Computational Tools for the *ab initio* Study of Electronic Properties of Crystals. *International Journal of Quantum Chemistry* **77**. 1032–1048.

Schaeffer H F III (Editor) 1977. *Applications of Electronic Structure Theory.* New York, Plenum Press.

Schaeffer H F III (Editor) 1977 *Methods of Electronic Structure Theory.* New York, Plenum Press.

Szabo A and N S Ostlund 1982. *Modern Quantum Chemistry Introduction to Advanced Electronic Structure Theory* New York, McGraw-Hill

Wimmer E 1991 Density Functional Theory for Solids, Surface and Molecules: from Energy Bands to Molecular Bonds In Labanowski J R and J W Andzelm (Editors) *Density Functional Methods in Chemistry.* Berlin, Springer-Verlag, pp 7–31

References

Almlöf J, K Faegri Jr and K Korsell 1982. Principles for a Direct SCF Approach to LCAO-MO *Ab initio* Calculations *Journal of Computational Chemistry* **3**.385–399

Ashcroft N W and N D Mermin 1976 *Solid State Physics* New York, Holt, Rinehart & Winston.

Baboul A G, L A Curtiss, P C Redfern and K Raghavachari 1999. Gaussian-3 Theory Using Density Functional Geometries and Zero-Point Energies. *Journal of Chemical Physics,* **110**:7650–7657.

Becke A D 1988. Density-functional Exchange-energy Approximation with Correct Asymptotic Behaviour. *Physical Review* **A38** 3098–3100.

Becke A D 1992 Density-functional Thermochemistry. I. The Effect of the Exchange-only Gradient Correction. *Journal of Chemical Physics* **96**:2155–2160

Becke A D 1993a. A New Mixing of Hartree-Fock and Local Density-functional Theories *Journal of Chemical Physics* **98** 1372–1377

Becke A D 1993b Density-functional Thermochemistry. III The Role of Exact Exchange *Journal of Chemical Physics* **98**.5648–5652.

Becke A D and R M Dickson 1990. Numerical Solution of the Schroedinger Equation in Polyatomic Molecules. *Journal of Chemical Physics* **92**:3610–3612

Bobrowicz F W and W A Goddard III 1977 The Self-Consistent Field Equations for Generalized Valence Bond and Open-Shell Hartree-Fock Wave Functions In Schaeffer H F III (Editor). *Modern Theoretical Chemistry III,* New York, Plenum, pp. 79–127.

Boys S F and F Bernardi 1970 The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies Some Procedures with Reduced Errors. *Molecular Physics* **19**:553–566

Bradley C J and A P Cracknell 1972 *The Mathematical Theory of Symmetry in Solids* Oxford, Clarendon Press.

Ceperley D M and B J Alder 1980 Ground State of the Electron Gas by a Stochastic Method. *Physical Review Letters* **45**:566–569.

Cohen M L 1986. Predicting New Solids and Superconductors. *Science* **234**:549–553

Cooper D L, J Gerratt and M Raimondi 1986. The Electronic Structure of the Benzene Molecule. *Nature* **323**. 699–701.

Curtiss L A, K Raghavachari, G W Trucks and J A Pople 1991 Gaussian-2 Theory for Molecular Energies of First- and Second-row Compounds. *Journal of Chemical Physics* **94**.7221–7230.

Curtiss L A, K Raghavachari, P C Redfern, V Rassolov and J A Pople 1998. Gaussian-3 (G3) Theory for Molecules Containing First and Second-row Atoms. *Journal of Chemical Physics* **109**:7764–7776

Curtiss L A, P C Redfern, K Raghavachari, V Rassolov and J A Pople 1999. Gaussian-3 Theory Using Reduced Møller–Plesset Order. *Journal of Chemical Physics* **110**:4703–4709

Dovesi R, R Orlando, C Roetti, C Pisani and V R Saunders 2000. The Periodic Hartree–Fock Method and Its Implementation in the CRYSTAL Code. *Physica Status Solidi* **B217**.63–88.

Dovesi R, C Pisani, C Roetti and V R Saunders 1983 Treatment of Coulomb Interactions in Hartree–Fock Calculations of Periodic-Systems *Physical Review* **B28**:5781–5792.

Foster J M and S F Boys 1960. Quantum Variational Calculations for a Range of CH_2 Configurations *Reviews in Modern Physics* **32**:305–307.

Frisch M J, G W Trucks and J R Cheeseman 1996. Systematic Model Chemistries Based on Density Functional Theory: Comparison with Traditional Models and with Experiment. *Theoretical and Computational Chemistry (Recent Developments and Applications of Modern Density Functional Theory)* 4:679–707.

Gerratt J, D L Cooper, P B Karadakov and M Raimondi 1997. Modern Valence Bond Theory. *Chemical Society Reviews* pp 87–100.

Gunnarsson O and B I Lundqvist 1976. Exchange and Correlation in Atoms, Molecules, and Solids by the Spin-density-functional Formalism. *Physical Review* **B13**:4274–4298.

Heine V 1970. The Pseudopotential Concept *Solid State Physics* **24**:1–36

Heitler W and F London 1927 Wechselwirkung neutraler Atome und Homöopolare Bindung nach der Quantenmechanik *Zeitschrift für Physik* **44**:455–472.

Hohenberg P and Kohn W 1964. Inhomogeneous Electron Gas. *Physical Review* **B136**:864–871.

Johnson B G, P M W Gill and J A Pople 1993. The performance of a family of density functional methods. *Journal of Chemical Physics* **98** 5612–5626

Kohn W and L J Sham 1965. Self-consistent Equations Including Exchange and Correlation Effects. *Physical Review* **A140**:1133–1138.

Lee C, W Yang and R G Parr 1988 Development of the Colle–Salvetti Correlation Energy Formula into a Functional of the Electron Density *Physical Review* **B37**:785–789.

Møller C and M S Plesset 1934. Note on an Approximate Treatment for Many-Electron Systems *Physical Review* **46**:618–622.

Monkhorst H J and J D Pack 1976 Special Points for Brillouin-zone Integration. *Physical Review* **B13**.5188–5192.

Morokuma K 1977. Why Do Molecules Interact? The Origin of Electron Donor–Acceptor Complexes, Hydrogen Bonding, and Proton Affinity. *Accounts of Chemical Research* **10** 294–300

Mujica A and R J Needs 1993. First-principles Calculations of the Structural Properties, Stability, and Band Structure of Complex Tetrahedral Phases of Germanium. ST12 and BC8. *Physical Review* **B48**:17010–17017.

Needs R J and Mujica 1995. First-principles Pseudopotential Study of the Structural Phases of Silicon. *Physical Review* **B51**:9652–9660.

Parr R G 1983 Density Functional Theory. *Annual Review of Physical Chemistry* **34**.631–656

Perdew J P and A Zunger 1981. Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems. *Physical Review* **B23**:5048–5079.

Pisani C and R Dovesi 1980. Exact-Exchange Hartree–Fock Calculations for Periodic Systems. I Illustration of the Method. *International Journal of Quantum Chemistry* **XVII**:501–516.

Pople J A, M Head-Gordon and K Raghavachari 1987. Quadratic Configuration Interaction A General Technique for Determining Electron Correlation Energies. *Journal of Chemical Phyics* **87**.5968–5975

Pople J A, M Head-Gordon, D J Fox, K Raghavachari and L A Curtiss 1989 Gaussian-1 Theory: A General Procedure for Prediction of Molecular Energies *Journal of Chemical Physics* **90**:5622–5629

Pople J A and R K Nesbet 1954. Self-consistent Orbitals for Radicals. *Journal of Chemical Physics* **22**.571–572

Pulay P 1977 Direct Use of the Gradient for Investigating Molecular Energy Surfaces In Schaeffer H F III (Editor). *Applications of Electronic Structure Theory*, New York, Plenum, pp 153–185.

Pulay P 1980. Convergence Acceleration of Iterative Sequences. The Case of SCF Iteration *Chemical Physics Letters* **73**:393–398.

Pulay P 1987 Analytical Derivative Methods in quantum Chemistry In Lawley K P (Editor) *Ab initio Methods in Quantum Chemistry – II*, New York, John Wiley & Sons, pp 241–286.

Roos B O, P R Taylor and E M Siegbahm 1980 A Complete Active Space SCF Method (CASSCF) Using a Density Matrix Formulated Super-CI Approach *Chemical Physics* **48** 157–173.

Schaeffer H F III 1986. Methylene: A Paradigm for Computational Quantum Chemistry *Science* **231**·1100–1107

Sim F, St-Amant A, I Papai and D R Salahub 1992. Gaussian Density Functional Calculations on Hydrogen-Bonded Systems *Journal of the American Chemical Society* **114**.4391–4400

Slater J C 1974. *Quantum Theory of Molecules and Solids Volume 4: The Self-Consistent Field for Molecules and Solids*. New York, McGraw-Hill

Smith B J, D J Swanton, J A Pople, H F Schaeffer III and L Radom 1990. Transition Structures for the Interchange of Hydrogen Atoms within the Water Dimer. *Journal of Chemical Physics* **92**·1240–1247.

St-Amant A, W D Cornell, P A Kollman and T A Halgren 1995. Calculation of Molecular Geometries, Relative Conformational Energies, Dipole Moments and Molecular Electrostatic Potential Fitted Charges of Small Organic Molecules of Biochemical Interest by Density Functional Theory. *Journal of Computational Chemistry* **16** 1483–1506

Stephens P J, F J Devlin, C F Chabalowski and M J Frisch 1994. *Ab Initio* Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *Journal of Physical Chemistry* **98**·11623–11627

Umeyama H and K Morokuma 1977. The Origin of Hydrogen Bonding An Energy Decomposition Study *Journal of the American Chemical Society* **99**·1316–1332

Vosko S H, L Wilk and M Nusair 1980. Accurate Spin-dependent Electron Liquid Correlation Energies for Local Spin Density Calculations· A Critical Analysis *Canadian Journal of Physics* **58**·1200–1211.

Wimmer E 1997. Electronic Structure Methods. In Catlow C R A and A K Cheetham (Editors). *New Trends in Materials Chemistry*, NATO ASI Series C 498 Dordrecht, Kluwer.