

accumulated (relative to unperturbed ions). In the lower panel, the same data is rendered using 3D isovalue surfaces. It is clear that the O ions at the step edge are polarised strongly. From this plot, the regions of strong polarisation can be quickly identified and examined in more detail.

## References

1. C. Pisani, R. Dovesi and C. Roetti, *Hartree-Fock ab initio Treatment of Crystalline Systems*, Lecture Notes in Chem. Vol. 48, Springer, Berlin (1988); R. Dovesi, C. Roetti, and V.R. Saunders, *CRYSTAL 92 User Documentation*, University of Torino, Torino (1992).
2. CETEP *user documentation* M.C. Payne, Cavendish Laboratory, University of Cambridge, UK.
3. R. Car and M. Parrinello, Phys Rev Lett **56**, 2471 (1985); M.P. Teter, M.C. Payne and D.C. Allan, Phys. Rev. B **40**, 12255 (1989); M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias and J.D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).
4. L.J. Clarke, I. Stich and M.C. Payne, Comput. Phys. Commun. **72**, 14 (1992).
5. DISPLAY *User documentation* Dr P Sherwood, Daresbury Laboratory, Daresbury, Warrington, WA4 4AD, UK.

## The CRYSTAL code

Carla Roetti

Department of Inorganic, Physical and Materials Chemistry, University of Torino, via Giuria 5, I-10125 Torino, Italy

May 28, 1996

**Summary.** The structure of the CRYSTAL code is analysed. This code permits the calculation of the wavefunction and properties of crystalline systems, using a Hartree-Fock Linear-Combination-of-Atomic-Orbitals (HF-LCAO) approximation. Both the theory and the practical implementation of the equations are discussed.

**Key words:** Hartree-Fock – Crystals – Coulomb Series – Exchange Series – Reciprocal Space Integration – Bipolar Approximation

## 1. Introduction

The CRYSTAL program permits the calculation of the wavefunction and properties of crystalline systems, within the Hartree-Fock Linear-Combination-of-Atomic-Orbitals (HF-LCAO) approximation.

The code was developed initially by the Theoretical Chemistry Group of the University of Torino at the end of the seventies [1, 2]. Since 1981 there has been a close collaboration on the project with V.R. Saunders (Daresbury Laboratory, UK) [3, 4]. The first public version was released in 1988 [5]. CRYSTAL92 [6] is now available as a FORTRAN source code, fully portable. The development of the current version of CRYSTAL has been greatly assisted by the constant support and interest of C. Pisani and by contributions from M. Causà, N. M. Harrison, E. Aprà and R. Orlando. The User's Manual includes 28 test cases, including both input and output, which provide examples of the wide range of program capabilities available.

A new release of the program is expected to appear in 1996. It is more efficient and has additional capabilities with respect to the current version, but the essential philosophy and techniques, which will be presented in this Chapter,

have remained unchanged. A discussion of the present scheme in relation to other approaches, was presented in Chapter 3. In the following sections, the equations are presented for the closed-shell case; extension to the open-shell case is discussed in Chapter 6.

## 2. Theoretical framework

CRYSTAL is an *ab initio* Hartree-Fock CO-LCAO program for the treatment of periodic systems.

CO-LCAO, in the present case, means that each Crystalline Orbital (CO),  $\psi_i(\mathbf{r}; \mathbf{k})$ , is a linear combination of Bloch functions (BF),  $\phi_\mu(\mathbf{r}; \mathbf{k})$ , defined in terms of local functions,  $\varphi_\mu(\mathbf{r})$  (here referred to as Atomic Orbitals, AOs).

$$\psi_i(\mathbf{r}; \mathbf{k}) = \sum_{\mu} a_{\mu,i}(\mathbf{k}) \phi_{\mu}(\mathbf{r}; \mathbf{k}) \quad (1)$$

$$\phi_{\mu}(\mathbf{r}; \mathbf{k}) = \sum_{\mathbf{T}} \varphi_{\mu}(\mathbf{r} - \mathbf{A}_{\mu} - \mathbf{T}) e^{i\mathbf{k} \cdot \mathbf{T}} \quad (2)$$

$\mathbf{A}_{\mu}$  denotes the coordinate of the nucleus in the zero reference cell on which  $\varphi_{\mu}$  is centred, and the  $\sum_{\mathbf{T}}$  is extended to the set of all lattice vectors,  $\mathbf{T}$ .

The local functions are expressed as linear combinations of a certain number,  $n_G$ , of individually normalized Gaussian type functions (GTO) characterized by the same centre, with fixed coefficients,  $d_j$  and exponents,  $\alpha_j$ , defined in the input:

$$\varphi_{\mu}(\mathbf{r} - \mathbf{A}_{\mu} - \mathbf{T}) = \sum_j^{n_G} d_j G(\alpha_j; \mathbf{r} - \mathbf{A}_{\mu} - \mathbf{T}) \quad (3)$$

The AOs belonging to a given atom are grouped into *shells*,  $\lambda$ . The shell can contain all AOs with the same quantum numbers,  $n$  and  $\ell$ , (for instance 3s, 2p, 3d shells), or all the AOs with the same principal quantum number,  $n$ , if the number of GTOs and the corresponding exponents are the same for all of them (mainly *sp* shells; this is known as the *sp constraint*). These groupings permit a reduction in the number of auxiliary functions that need to be calculated in the evaluation of electron integrals and therefore, increase the speed of calculation.

A single, normalized, *s*-type GTO,  $G_{\lambda}$ , is associated with each shell (the *adjoined Gaussian* of shell  $\lambda$ ). The  $\alpha$  exponent is the smallest of the  $\alpha_j$  exponents of the Gaussians in the contraction. The adjoined Gaussian is used to estimate the AO overlap and select the level of approximation to be adopted for the evaluation of the integrals.

The expansion coefficients of the Bloch functions,  $a_{\mu,i}(\mathbf{k})$ , are calculated by solving the matrix equation for each reciprocal lattice vector,  $\mathbf{k}$ :

$$\mathbf{F}(\mathbf{k})\mathbf{A}(\mathbf{k}) = \mathbf{S}(\mathbf{k})\mathbf{A}(\mathbf{k})\mathbf{E}(\mathbf{k}) \quad (4)$$

in which  $\mathbf{S}(\mathbf{k})$  is the overlap matrix over the Bloch functions,  $\mathbf{E}(\mathbf{k})$  is the diagonal energy matrix and  $\mathbf{F}(\mathbf{k})$  is the Fock matrix in reciprocal space:

$$\mathbf{F}(\mathbf{k}) = \sum_{\mathbf{T}} \mathbf{F}^{\mathbf{T}} e^{i\mathbf{k} \cdot \mathbf{T}} \quad (5)$$

The matrix elements of  $\mathbf{F}^{\mathbf{T}}$ , the Fock matrix in direct space, can be written as a sum of one-electron and two-electron contributions in the basis set of the AO:

$$F_{12}^{\mathbf{T}} = H_{12}^{\mathbf{T}} + B_{12}^{\mathbf{T}} \quad (6)$$

The one-electron contribution is the sum of the kinetic and nuclear attraction terms:

$$H_{12}^{\mathbf{T}} = T_{12}^{\mathbf{T}} + Z_{12}^{\mathbf{T}} = \langle \varphi_1^0 | \hat{T} | \varphi_2^{\mathbf{T}} \rangle + \langle \varphi_1^0 | \hat{Z} | \varphi_2^{\mathbf{T}} \rangle \quad (7)$$

In core pseudopotential calculations,  $\hat{Z}$  includes the sum of the atomic pseudopotentials.

The two-electron term is the sum of the Coulomb and exchange contributions:

$$B_{12}^{\mathbf{T}} = C_{12}^{\mathbf{T}} + X_{12}^{\mathbf{T}} = \quad (8)$$

$$\sum_{3,4} \sum_{\mathbf{Q}} P_{3,4}^{\mathbf{Q}} \sum_{\mathbf{S}} [(\varphi_1^0 \varphi_2^{\mathbf{T}} | \varphi_3^{\mathbf{S}} \varphi_4^{\mathbf{S}+\mathbf{Q}}) - \frac{1}{2}(\varphi_1^0 \varphi_3^{\mathbf{S}} | \varphi_2^{\mathbf{T}} \varphi_4^{\mathbf{S}+\mathbf{Q}})] \quad (9)$$

The Coulomb interactions, that is, those of electron-nucleus, electron-electron and nucleus-nucleus, are individually divergent, due to the infinite size of the system. The grouping of corresponding terms is necessary in order to eliminate this divergence.

The  $P^{\mathbf{Q}}$  density matrix elements in the AOs basis set are computed by integration over the volume of the Brillouin zone (BZ),

$$P_{3,4}^{\mathbf{Q}} = 2 \int_{BZ} d\mathbf{k} e^{i\mathbf{k} \cdot \mathbf{Q}} \sum_n a_{3n}^*(\mathbf{k}) a_{4n}(\mathbf{k}) \theta(\epsilon_F - \epsilon_n(\mathbf{k})) \quad (10)$$

where  $a_{in}$  denotes the  $i$ -th component of the  $n$ -th eigenvector,  $\theta$  is the step function,  $\epsilon_F$ , the Fermi energy and  $\epsilon_n$ , the  $n$ -th eigenvalue.

The total electronic energy per unit cell is given by:

$$E^{elec} = \frac{1}{2} \sum_{1,2} \sum_{\mathbf{T}} P_{12}^{\mathbf{T}} (H_{12}^{\mathbf{T}} + B_{12}^{\mathbf{T}}) \quad (11)$$

Chapter 11 presents a discussion of the different contributions to the total energy.

### 2.1. Remarks on the evaluation of the integrals

The approach adopted for the treatment of the Coulomb and exchange series is based on a few simple ideas and on a few general tools, which can be summarized as follows:

1. Where possible, terms of the Coulomb series are aggregated so as to reduce the number of integrals to be evaluated;
2. Exchange integrals which will combine with small density matrix elements are disregarded;
3. Integrals between non-overlapping distributions are approximated;
4. Approximations for large integrals must be very accurate; for small integrals large percentage errors can be accepted;
5. Selection must be very efficient, because a large number of possible terms must be checked (adjoined Gaussians are very useful from this point of view);

## 2.2. Treatment of the Coulomb series

For the evaluation of the Coulomb contributions to the total energy and Fock matrix, correct coupling of electron-nucleus and electron-electron interactions is essential. The computational technique for doing so was presented by Dovesi *et al.* [3] and by Saunders *et al.* [7]. It may be summarized as follows.

Consider the Coulomb bielectronic contribution to the Fock matrix ( $C_{12}^T$ ) and to the total energy :

$$E_{ee}^{coul} = \frac{1}{2} \sum_{1,2} \sum_T P_{12}^T \sum_{3,4} \sum_Q P_{34}^Q \sum_S [(\varphi_1^0 \varphi_2^T | \varphi_3^S \varphi_4^{S+Q})] \quad (12)$$

Seven indices are involved in equation 12; four of them (1, 2, 3 and 4) refer to the AOs of the unit cell; in principle, the other three (T, Q and S) span the infinite set of translation vectors: for example,  $\varphi_2^T(\mathbf{r})$  is AO number 2 in cell T.  $P$  is the density matrix; the usual notation is used for the bielectronic integrals. Due to the localized nature of the basis set, the total charges,  $q_1$  and  $q_2$ , associated with the two overlap distributions:  $\{G_{10} G_{2T}\}$  and  $\{G_{3S} G_{4S+Q}\}$ , decay exponentially to zero with increasing  $|T|$  and  $|Q|$  (for example,  $G_1$  is the adjoined Gaussian of the shell to which  $\varphi_1$  belongs).

A Coulomb overlap parameter,  $S_c$ , can be defined in such a way that when either  $q_1$  or  $q_2$  are smaller than  $S_c$ , the bielectronic integral is disregarded, and the sum over T or Q truncated. The ITOL1 input parameter is defined as  $ITOL1 = -\log_{10} S_c$ . The same parameter value is used for selecting overlap, kinetic and multipole integrals.

The problem of the S summation in equation 12 is more delicate, S being related to the distance between the two interacting distributions. The multipolar expansion scheme illustrated below is particularly effective when large unit cell or low dimensionality systems are considered.

The electron-electron and electron-nuclei series ( $C_{12}^T$  and  $Z_{12}^T$ ) can be rearranged as follows:

1. Mulliken shell *net* charge distributions are defined as :

$$\rho_\lambda(\mathbf{r}-\mathbf{S}) \equiv \{\lambda\}' \equiv \{\lambda\}^* - Z_\lambda = \sum_{3 \in \lambda} \sum_Q P_{34}^Q \varphi_3(\mathbf{r}-\mathbf{S}) \varphi_4(\mathbf{r}-\mathbf{S}-\mathbf{Q}) - Z_\lambda \quad (13)$$

where  $Z_\lambda$  is the fraction of nuclear charge formally attributed to shell  $\lambda$ , and  $\{\lambda\}$  is the electronic charge distribution of shell  $\lambda$ .

2. Z and C contributions are reordered:

$$C_{12}^T + Z_{12}^T = \sum_\lambda \sum_S \int d\mathbf{r} d\mathbf{r}' \varphi_1^0(\mathbf{r}) \varphi_2^T(\mathbf{r}) |\mathbf{r}-\mathbf{r}'-\mathbf{S}|^{-1} \rho_\lambda(\mathbf{r}'-\mathbf{S}) \quad (14)$$

3. For a given shell  $\lambda$ , there is a finite set of S vectors,  $B_\lambda$ , which the two interacting distributions overlap; in this  $B_\lambda$  zone (*bielectronic zone*), all the bielectronic integrals are evaluated explicitly. In the outer, infinite region which we define as  $M_\lambda$ , complementary to  $B_\lambda$  (the *monoelectronic zone*),  $\rho_\lambda$  can be expanded in multipoles and the series can be evaluated to infinity analytically, using Ewald's method combined with recursion formulae [7].

The resulting expression for the Coulomb contribution to the Fock matrix is:

$$\begin{aligned} C_{12}^T + Z_{12}^T &= \sum_\lambda \left\{ \sum_S \left[ \sum_{3 \in \lambda} \sum_Q P_{34}^Q (\varphi_1^0 \varphi_2^T | \varphi_3^S \varphi_4^{S+Q}) \right. \right. \\ &\quad \left. \left. - \sum_{\ell, m} \gamma_\ell^m(\mathbf{A}_\lambda; \{\lambda\}) \Phi_\ell^m(12T; \mathbf{A}_\lambda + \mathbf{S}) \right] + \right. \\ &\quad \left. + \sum_S \sum_{\ell, m} \gamma_\ell^m(\mathbf{A}_\lambda; \{\lambda\}') \Phi_\ell^m(12T; \mathbf{A}_\lambda + \mathbf{S}) \right\} \end{aligned} \quad (15)$$

where:

$$\gamma_\ell^m(\mathbf{A}_\lambda; \{\lambda\}) = \int d\mathbf{r} \rho_\lambda(\mathbf{r}-\mathbf{A}_\lambda) N_\ell^m X_\ell^m(\mathbf{r}-\mathbf{A}_\lambda) \quad \text{and} \quad (16)$$

$$\Phi_\ell^m(12T; \mathbf{A}_\lambda + \mathbf{S}) = \int d\mathbf{r} \varphi_1^0(\mathbf{r}) \varphi_2^T(\mathbf{r}) X_\ell^m(\mathbf{r}-\mathbf{A}_\lambda - \mathbf{S}) |\mathbf{r}-\mathbf{A}_\lambda - \mathbf{S}|^{-2\ell-1} \quad (17)$$

The  $X_\ell^m$  functions entering in the definition of the multipoles and field terms are real, solid harmonics and  $N_\ell^m$ , the corresponding normalization coefficients.

The advantage of using equation 15 is that many four-centre (long-range) integrals can be replaced by a smaller number of three-centre integrals.

The attribution of the interaction between  $\rho_1 = \{10 2T\}$  and  $\rho_\lambda$  either to the *exact*, short-range or to the *approximate*, long-range zone is performed by comparing the penetration between  $\rho_1$  and  $\rho_\lambda$  with the ITOL2 input parameter (if  $ITOL2 > -\log S_{1\lambda}$ , then  $\rho_\lambda$  is attributed to the *exact*  $B_\lambda$  zone).

The multipolar expansion, in the approximate zone, is truncated at  $L = \ell^{max}$ . The default value of  $L$  is 4; the maximum possible value is 6, the minimum suggested value, 2.

## 2.3. The exchange series

The exchange series does not require particular manipulations of the kind discussed in the previous subsection, but needs a careful selection of the terms contributing appreciably to the Fock operator and to the total energy [4]. The exchange contribution to the total energy can be written as follows:

$$E^{ex} = \frac{1}{2} \sum_{12} \sum_T P_{12}^T \left[ -\frac{1}{2} \sum_{34} \sum_Q P_{34}^Q \sum_S (\varphi_1^0 \varphi_3^S | \varphi_2^T \varphi_4^{S+Q}) \right] \quad (18)$$

where the term in square brackets is the exchange contribution to the 12T element of the direct-space Fock matrix.  $E^{ex}$  has no counterpart of opposite sign as the Coulomb term has; hence, it must converge by itself.

The S summation can be truncated after a few terms, since the  $\{\varphi_1^0 \varphi_3^S\}$  overlap distribution decays exponentially as S increases. Similar considerations apply to the second charge distribution. In CRYSTAL, the S summation is, therefore, truncated when the charge associated with either  $\{G_{10} G_{3S}\}$  or  $\{G_{2T} G_{4S+Q}\}$  is smaller than  $10^{-ITOL3}$ .

The situation is more complicated when T and Q summations are analysed. Let us consider the leading terms at large distance, corresponding to  $\varphi_1 = \varphi_3$ ,  $\varphi_2 = \varphi_4$ ,  $\mathbf{S} = \mathbf{0}$  and  $\mathbf{Q} = \mathbf{T}$ :

$$e_{12}^T = -1/4(P_{12}^T)^2(10\ 10|2T\ 2T) = -(p^T)^2/(4|T|) \quad (19)$$

(Here,  $p^T$  indicates the dominant  $P$  matrix element at long-range). Since the number of terms per unit distance of this kind increases as  $|T|^{d-1}$ , where  $d$  is the dimensionality of the system, it is clear that the convergence of the series depends critically on the long-range behaviour of the bond order matrix.

Cancellation effects, associated in particular with the oscillatory behaviour of the density matrix in metallic systems, are not predominant at long-range. Even if the actual behaviour of the  $P$  matrix elements cannot be predicted because it depends, in a complicated way, on the physical nature of the compound [8], on orthogonality constraints and on basis set quality, the different range of valence and core elements can be exploited by adopting a *pseudoverlap* criterion. This consists in truncating  $T$  summations when the  $\int dr \varphi_1^0 \varphi_2^T$  overlap is smaller than a given threshold, defined as  $P_{ex}^T$  (where  $ITOL4 = -\log_{10}(P_{ex}^T)$ ) and also truncating the  $Q$  summation when  $\int dr \varphi_3^0 \varphi_4^Q$  overlap is smaller than the threshold,  $P_{ex}^Q$  ( $ITOL5 = -\log_{10}(P_{ex}^Q)$ ).

Despite its partially arbitrary nature, this criterion presents some advantages with respect to other, more elaborate schemes: it is similar to the other truncation schemes ( $ITOL1$ ,  $ITOL2$  and  $ITOL3$ ) and so the same classification tables can be used; it is, in addition, reasonably efficient in terms of space occupation and computer time.

This truncation scheme is symmetrical with respect to the  $T$  and  $Q$  summations. However, if account is not taken of the different role of the two summations in the SC (Self-Consistent) stage, distortions may be generated in the exchange field, as felt by charge distributions  $\varphi_1 \varphi_2^T$ , where  $T$  labels the largest (in modulus)  $T$  vector taken into account according to  $ITOL4$ . This distortion may be variationally *exploited* and unphysically large density matrix elements build up progressively along the SC stage, eventually leading to catastrophic behaviour (see Chapter II.5 of reference [2] for a discussion of this point). In order to overcome this problem, the threshold,  $P_{ex}^Q$  ( $ITOL5$ ) for  $Q$  summation must be more severe than that for  $T$  summation ( $ITOL4$ ). In this way, all the integrals whose second pseudocharge  $\int dr \varphi_3^0 \varphi_4^Q$  is larger than  $P_{ex}^Q$  are taken into account. A difference in the two thresholds ranging from three to eight orders of magnitude is sufficient to stabilize the SC behaviour in most cases.

#### 2.4. Bipolar expansion approximation of Coulomb and exchange integrals

We may now return to the partition of the  $S$  summation in the Coulomb series, shown in equation 12. Consider one contribution to the charge distribution of electron 1, centred in the reference cell:  $\rho^0 = \varphi_1^0 \varphi_2^T$ ; now consider the charge distribution  $\rho_\lambda(S)$  of shell  $\lambda$ , centred in cell  $S$  (equation 13). For small  $|S|$  values,  $\rho_\lambda$  and  $\rho^0$  overlap, so that all the related bielectronic integrals must be evaluated exactly, one by one; for larger values of  $|S|$ ,  $\rho_\lambda$  is external to  $\rho^0$ , so that all the related bielectronic integrals are grouped and evaluated, in an approximate way, through the multipolar expansion of  $\rho_\lambda$ .

However, in many instances, although  $\rho_\lambda$  is not external to  $\rho^0$ , the two-centre  $\varphi_3^S \varphi_4^{S+Q}$  contributions to  $\rho_\lambda$  are external to  $\rho^0 = \varphi_1^0 \varphi_2^T$ ; in this case, instead of exactly evaluating the bielectronic integral, a two-centre, truncated, bipolar expansion can be used (see Chapter II.4.c in reference [2] and references therein).

In order to decide to which zone a shell may be ascribed, we proceed as follows: when, for a given pair of shells  $\lambda_1^0 \lambda_2^T$ , shell  $\lambda_3^S$  is attributed to the  $B$  (bielectronic) zone, the penetration between the products of adjoined Gaussians  $G_1^0 G_2^T$  and  $G_3^S G_4^{S+Q}$  is estimated: the default value of the penetration parameter is 14, and the block of bielectronic integrals is attributed accordingly to the  $b_e$  (exact) or to the  $b_b$  (bipolar) zone. The set of  $S$  vectors, defining the  $B$  zone of  $\rho^0 = \{12T\}$  and  $\rho_\lambda \equiv \{\lambda_3\}$ , is then split into two subsets, which are specific for each partner  $\lambda_4^i$  of  $\lambda_3$ .

A similar scheme is adopted for the selected exchange integrals (see previous section) whose pseudocharges do not overlap appreciably. The default value of the penetration parameter is 10.

The total energy change due to the bipolar expansion approximation should not be greater than  $10^{-4}$  Hartree/atom; exact evaluation of all the two-electron integrals (obtained by setting the penetration parameter value  $> 20000$ ) increases the computational cost by a factor of between 1.3 and 3. Multipolar expansion is very efficient, because the following two conditions are fulfilled:

1. A general algorithm is available for reaching high  $\ell$  values easily and economically [3, 7]. The maximum allowed value is  $\ell=6$ .
2. The multipolar series converges rapidly, either because the interacting distributions are nearly spherical (shell expansion) or because their functional expression is such that their multipoles are zero above a certain (low)  $\ell$  value.

#### 2.5. Exploitation of symmetry

Translational symmetry allows the factorization of the eigenvalue problem in periodic calculations, because the Bloch functions are a basis for irreducible representations of the translational group.

In periodic calculations, point symmetry is exploited to reduce the number of points for which the matrix equations are to be solved. Point symmetry is also explicitly used in the reconstruction of the Hamiltonian, which is totally symmetric with respect to the point group operators of the system.

In the HF-CO-LCAO scheme, the very extensive use of point symmetry allows us to evaluate bielectronic and monolectronic integrals with saving factors as large as  $h$ , in the number of bielectronic integrals to be computed or  $h^2$ , in the number of those to be stored for the SCF part of the calculation, where  $h$  is the order of the point group. The main steps of the procedure [11] can be summarized as follows:

- The set of Coulomb and exchange integrals, whose indices 3 and 4 (3, 4) identify translationally equivalent pairs of AOs, so that the associated element of the density matrix  $P_{34}$  is the same, are summed together to give  $D_{1234}$  elements:

$$D_{1,2T;3,4Q} = \sum_Q [(\varphi_1^0 \varphi_2^T | \varphi_3^S \varphi_4^{S+Q}) - 1/2(\varphi_1^0 \varphi_3^Q | \varphi_2^T \varphi_4^{S+Q})] \quad (20)$$

- The products of AOs  $\varphi_1 \varphi_2$  (and  $\varphi_3 \varphi_4$ ) are classified in symmetry-related sets; using the fact that the Fock matrix is totally symmetric, only those quantities are evaluated whose indices 1 and 2 (1, 2) refer to the first member of a symmetry set. The corresponding saving factor is as large as  $h$ .

- Using the symmetry properties of the density matrix, DM, quantities referring to 3, 4 couples belonging to the same symmetry set (and with the same 1, 2T index) can be combined, after multiplication by appropriate symmetry matrices, so that a single quantity for each 3, 4 symmetry set is to be stored, with a saving factor in storage of the order of  $h$ .
- The symmetry  $P_{34}^Q = P_{43}^{-Q}$  is exploited.
- The symmetry  $F_{12}^T = F_{21}^{-T}$  is exploited.

## 2.6. Reciprocal space integration

The integration in reciprocal space is an important aspect of *ab initio* calculations for periodic structures. The problem arises at each stage of the self-consistent procedure, when determining the Fermi energy,  $\epsilon_F$ , when reconstructing the one-electron density matrix and, after self-consistency is reached, when calculating the density of states (DOS) and a number of observable quantities. The  $P$  matrix in direct space is computed following equation 10. The technique adopted to compute  $\epsilon_F$  and the  $P$  matrix in the SCF step is described in reference [1]. The Fourier-Legendre technique presented in Chapter II.6 of reference [2] is adopted in the calculation of total and projected DOS. The Fermi energy and the integral in equation 10 are evaluated starting from the knowledge of the eigenvalues,  $\epsilon_n(\kappa)$  and the eigenvectors,  $a_{\mu n}(\kappa)$ , at a certain set of sampling points,  $\{\kappa\}$ . In 3D crystals, the sampling points belong to a lattice (called the *Monkhorst net*, [9]) with basis vectors  $b_1/s_1$ ,  $b_2/s_2$ ,  $b_3/s_3$ , where  $b_1$ ,  $b_2$  and  $b_3$  are the ordinary reciprocal lattice vectors;  $s_1$ ,  $s_2$  and  $s_3$  (input as IS1, IS2 and IS3) are integer *shrinking factors*. Unless otherwise specified, IS1=IS2=IS3=IS. In 2D crystals, IS3 is set equal to 1; in 1D crystals both IS2 and IS3 are set equal to 1. Only points of the Monkhorst net belonging to the irreducible part of the Brillouin Zone (BZ) are considered, having associated geometrical weights,  $w_i$ .

In the selection of the  $\kappa$  points for non-centrosymmetric crystals, time-reversal symmetry is exploited ( $\epsilon_n(\kappa) = \epsilon_n(-\kappa)$ ).

The number of inequivalent sampling points,  $\kappa_i$ , is given asymptotically by the product of the shrinking factors divided by the order of the point group. In high symmetry systems and with small  $s_i$  values, it may be considerably larger because many points lie on symmetry planes or axes.

Two completely different situations (which are automatically identified by the code) must now be considered, depending on whether the system is an insulator (or zero-gap semiconductor) or a conductor. In the former case, all bands are either fully occupied or vacant. The identification of  $\epsilon_F$  is elementary and the Fourier transform expressed by equation 10 is reduced to a weighted sum of the integrand function over the set  $\{\kappa_i\}$  with weights  $w_i$ , the sum over  $n$  being limited to occupied bands.

The case of conductors is more complicated: two additional parameters, ISHF and ISP, enter into play. ISHF is the number of symmetrized plane waves,  $A_m(\mathbf{k})$ , used for representing the  $\mathbf{k}$  dependence of the eigenvalues. ISP (or ISP1, ISP2, ISP3) are *Gilat shrinking factors* which define a *Gilat net* [10], completely analogous to the Monkhorst net. The value of ISP is larger than IS (by up to a factor of 2), giving a denser net. Each point,  $\kappa'_j$ , has a small sphere,  $S_j$ , attached, centred at  $\kappa'_j$  and with volume  $v = V_{BZ}/(s_1 s_2 s_3)$ , that is, equal to the inverse

density of Gilat points: the spheres are, therefore, partially overlapping. In a conductor, one or more bands are partially occupied. Suppose, for simplicity, that there is only one such band (the  $n$ -th, say). A truncated Fourier expansion of  $\epsilon_n(\mathbf{k})$  is considered:

$$\epsilon_n(\mathbf{k}) \approx \tilde{\epsilon}_n(\mathbf{k}) = \sum_{m=1}^{ISHF} \tilde{\epsilon}_n A_m(\mathbf{k}) \quad (21)$$

with expansion coefficients,  $\tilde{\epsilon}_n$ , obtained from the knowledge of  $\epsilon_n(\kappa)$  at the Monkhorst sampling points. We next consider a further approximation:  $\tilde{\epsilon}_n(\mathbf{k})$  of  $\epsilon_n(\mathbf{k})$ , corresponding to linearly expanding  $\tilde{\epsilon}_n(\mathbf{k})$  around each Gilat point  $\kappa_j$  within its own sphere,  $S_j$ . Using this approximation for  $\epsilon_n(\mathbf{k})$ , it is possible to obtain  $\epsilon_F$  easily and then to define appropriate weights,  $w'_{in}$ , for each Monkhorst point and each band. The integral [10] is expressed, as before, as a weighted sum, but with weights  $w'_{in}$  instead of  $w_i$ . The accuracy of this procedure increases with increasing IS, ISP and ISHF.

In high symmetry systems, it is convenient to assign IS *magic* values such that all low multiplicity (high symmetry) points belong to the Monkhorst lattice. Although this choice does not correspond to maximum efficiency, it gives a safer estimate of the integral.

The values assigned to ISHF and ISP are irrelevant for non-conductors. However, a non-conductor may give rise to a conducting structure at the initial stages of the SCF cycle, owing, for instance, to a very unbalanced initial guess at the density matrix. The ISHF and ISP parameters must therefore be defined in all cases.

## 2.7. Basis set selection

The problem of selecting appropriate basis sets of contracted GTOs, for HF calculations of crystals, has many features in common with the corresponding one for molecules; yet, some aspects are markedly different in the two cases.

Diffuse Gaussian orbitals (exponent coefficient  $< 0.1$ ) play a critical role in HF-CO-LCAO calculations of periodic structures, especially three-dimensional ones.

The number of integrals to be calculated increases dramatically with decreasing exponent. The risk of pseudo-linear dependence increases rapidly, demanding higher precision in order to avoid catastrophic behaviour. On the other hand, very diffuse AOs are much less important in three-dimensional, densely packed crystals than in atoms or molecules, where they serve to describe the tails of the electronic distribution towards the vacuum. Another effect of the much higher density of atoms in crystals, with respect to molecules, is that relatively poor basis sets provide larger variational freedom in the former case, since valence orbitals are, in fact, shared among a large number of neighbours and reaching the HF limit becomes easier.

Problems may arise when the binding energy is estimated by subtracting the total crystal energy per unit cell from the sum of HF energies of the isolated atoms. Using the same AOs for the isolated atoms as are used in the crystal

necessarily overestimates the HF binding energy (Basis Set Superposition Error, BSSE). In order to allow for this, counterpoise corrections may be used.

As a general rule, atomic basis sets optimized for molecules can be the best starting point to build a basis set for an atom in a given crystal and in a given chemical bonding scheme. For instance, three different basis sets must be used (and optimized) for oxygen in MgO and H<sub>2</sub>O (ice) [12, 13].

## 2.8. Calculation of observable quantities

The total electronic energy is the fundamental observable. However, comparing energy data requires some care. It is customary to extrapolate formation energies to zero temperature, to subtract nuclear-zero-point-motion effects and to add isolated atom energies to obtain values which may be compared with the experimental total electronic energy of the crystals. This procedure implicitly assumes the validity of the Born-Oppenheimer separation of nuclear and electronic motion. Such an approximation is a relatively minor one, when compared with the other errors that affect calculated HF energies, due to basis set incompleteness and numerical approximations. These errors strongly affect the absolute value of the total energy, but they cancel, to a large extent, when the energies of systems containing the same atoms in different configurations are compared.

The Hartree-Fock energy can be corrected for the neglect of correlation terms *a posteriori*, through a functional of the charge density [14, 15].

CRYSTAL can compute the following quantities, from a HF-CO-LCAO SCF wavefunction:

- Band structure; density of states
- Electronic charge density; charge density gradient; x-ray structure factors
- Atomic multipole and spheropole; Mulliken population analysis; topological analysis of charge density [16]
- Electron momentum distribution; Compton profiles and related quantities
- Electrostatic potential; electric field and electric field gradient

## 3. Structure of the code\*

The CRYSTAL code consists of a suite of three programs:

**integrals**, **SCF** and **properties**. The current version allows direct SCF execution.

The structure of the code closely follows the theory and is similar to those of molecular programs. In the flow diagram on the following page, the peculiar features of periodic Hartree-Fock are in bold-face.

Program communication is provided by means of a common filing system. Unformatted sequential data sets are used. Interfaces to molecular graphics programs and quantum molecular programs are supplied.

The *s*, *p* (in the order *x*, *y* and *z*) and *d* (in the order  $2z^2 - x^2 - y^2$ , *xz*, *yz*,  $x^2 - y^2$ , *xy*) shells of GTF can be used. Shells of the type *sp* are also available (*s* and *p* shells, sharing the same set of exponents). The use of *sp* shells can give rise to considerable savings in CPU time.

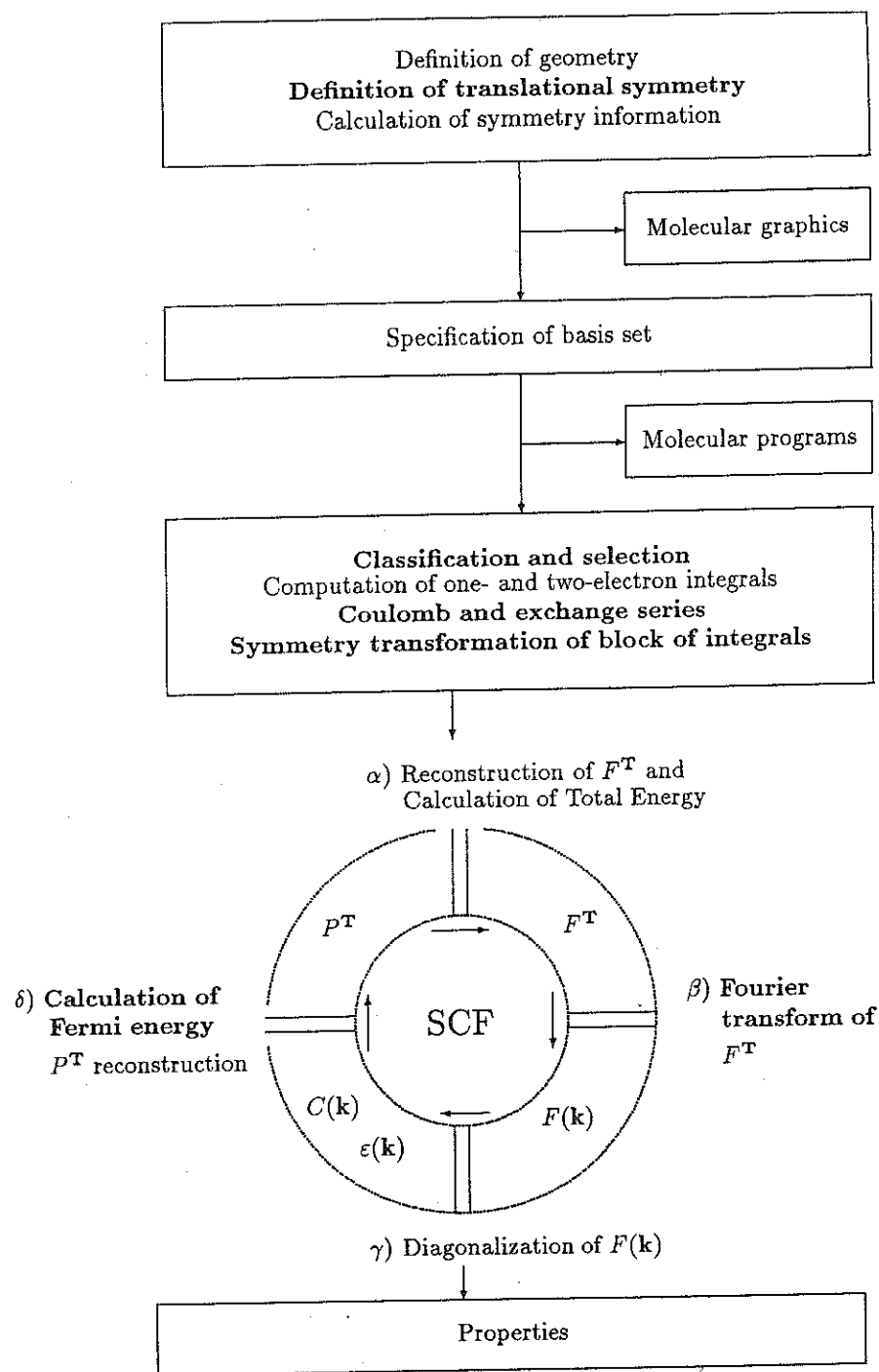


Figure 1. Flow chart of the CRYSTAL code.

The program can automatically handle space symmetry (230 space groups, 80 two-sided plane groups, 99 line groups, 45 point groups are available). In the case of polymers, it cannot treat helical structures (translation followed by a rotation around the periodic axis). However, when commensurate rotations are involved, a suitably large unit cell can be adopted. Point symmetries *compatible with translation symmetry* are provided for molecules.

The program can perform closed-shell, restricted Hartree-Fock, Restricted Open Shell (ROHF) and Unrestricted Open shell (UHF) calculations.

All-electron and valence-only basis sets with core pseudo-potentials are allowed.

### 3.1. Limits of applicability

The code is written in FORTRAN, standard ANSI77 and it has been tested on a large number of computer systems, from PCs to work-stations and CRAYs. It is readily portable. The distribution tape contains a makefile for Unix systems, to generate the executable files. No external libraries are required.

All the key dimensions are parameterized; in principle, it should be possible to run a system with as many atoms and basis functions as allowed by numerical accuracy, with good computational parameters. However, there are many dimensions which are very difficult to predict, as they depend in a correlated way on the number of symmetry-related atoms, the size of the basis set, the number of points in the reciprocal lattice used to compute Fermi energy and the required accuracy as specified by various tolerances. There are many checks on the dimensions of the system and suggestions are printed before the program stops, if such checks indicate an error condition. It is, therefore, difficult to define, in a few words, the maximum size of the systems that can be handled by CRYSTAL (see [6], Appendix F).

## 4. The cost of the wavefunction calculation

The cost of the calculation depends on many factors. Some of them can be controlled by adjusting the computational parameters, according to the precision required in the results. Information, such as formation energy, charges and band structure, do not require severe conditions. When looking for phase transitions or calculating the elastic properties, more severe conditions are required as total energy derivatives are computed numerically (the analytical gradient code is not yet implemented).

The cost of the **integrals** step is related to the number of integrals evaluated exactly and depends, therefore, on the following parameters:

- **The number of basis functions in the unit cell.**

Given a basis set, the number of two-electron integrals evaluated exactly, depends almost linearly on the cell size, while the number of monoelectronic integrals has a quadratic dependence.

- **Symmetry of the system.**

In high symmetry systems only a small fraction of integrals are actually evaluated, all the others being generated by symmetry transformations.

- **Exponents of the valence Gaussians.**

The selection criteria of the integrals to be computed is based on the overlap of the lowest exponent Gaussians. Avoid very diffuse functions (exponents  $0.1 \text{ Bohr}^{-2}$  or less).

- **Type of basis functions.**

*d* orbitals are much more expensive to use than *s* and *p* orbitals; *sp* shells are much less expensive than split *s* and *p* shells.

The disk space to be allocated can be estimated by using a test-run option; the SCF direct program eliminates the storage of the bielectronic integrals that are computed at each SC cycle.

In the SCF step, most of the time is spent in matrix algebra operations, such as matrix multiplication and matrix diagonalization, which scale as the third power of the number of basis functions, when traditional numerical techniques are adopted. SCF time is proportional to i) the number of SCF cycles; ii) the number of reciprocal lattice points,  $\kappa$ , at which the Fock matrix is diagonalized (for the calculation of Fermi energy and the density matrix).