CRYSCOR09

C. Pisani[(1)], S. Casassa[(1)], L. Maschio[(1)]
M. Schütz[(2)], D. Usvyat[(2)]

(1) Theoretical Chemistry Group, Università di Torino - Torino (Italy)

(2) Institute for Physical and Theoretical Chemistry, Universität Regensburg - Regensburg (Germany)

User's Manual

A. Erba[(1)], M. Halo[(1)]

March 5, 2010

Contents

1 Generalities **5**
1.1 Introduction . . . 5
1.2 Conventions . . . 5
1.3 Functionality . . . 6
1.4 Citing CRYSCOR . . . 7
1.5 The MOLPRO contribution . . . 7
1.6 How to run a CRYSCOR job . . . 7
1.7 Bugs Reporting . . . 8
1.8 Acknowledgments . . . 8

2 The main ingredient: Wannier Functions **9**
2.1 The symmetrization step . . . 10
2.1.1 An example: the three-layers slab of MgO . . . 10

3 Input format and mandatory keywords **14**
3.1 Input format . . . 14
3.2 General and mandatory keywords . . . 16

4 Occupied and virtual space description **17**
4.1 Occupied space description . . . 17
4.1.1 Definition of kernels . . . 17
4.1.2 Pair selection . . . 18
4.1.3 Lennard-Jones extrapolation . . . 18
4.1.4 Pair partitioning in adsorption problems . . . 19
4.2 Virtual space description . . . 20
4.2.1 (Excitation) Domains: General definitions . . . 20
4.2.2 Domain definition: Mulliken Population Analysis . . . 21
4.2.3 Domain definition: Stars of neighbors . . . 21
4.2.4 Domain definition: Molecular domains . . . 22
4.2.5 Domain definition: Coefficients of WFs . . . 22
4.2.6 Checking the shape of domains . . . 22
4.3 Truncation criteria . . . 23
4.4 Fixing of indices for manual geometry optimization . . . 24

5 LMP2 integrals and equations **25**
5.1 Two-electron integrals . . . 25
5.1.1 Exact two-electron integrals . . . 25
5.1.2 Multipolar Expansion . . . 26
5.1.3 Density Fitting . . . 26
5.2 MP2 Equations . . . 30
5.3 Restart keywords . . . 30

6 Dual Basis Set **31**

7 LMP2 correction to the HF one-electron Density Matrix **32**

8 Printing and Plotting **34**
8.1 Timing information . . . 36

9 Possible sources of errors **37**

10 What the output looks like **39**
10.1 The simple case of the LiH crystal . . . 39
10.1.1 The Header and the recovery of the information from CRYSTAL . . . 40
10.1.2 The recovery of the information from PROPERTIES . . . 42
10.1.3 The computational setup of the MP2 calculation . . . 42
10.1.4 The operating region . . . 43
10.1.5 The excitations domains . . . 44
10.1.6 The kernel domains . . . 44
10.1.7 The pairs partition (distance) . . . 45
10.1.8 Two-electrons Integrals . . . 46
10.1.9 LMP2 equations . . . 47
10.1.10 The Grimme correction . . . 47
10.1.11 Singly excited configurations . . . 48
10.1.12 Final Summary . . . 48
10.1.13 The Lennard-Jones extrapolation to infinity . . . 48
10.1.14 MP2 correction to the HF Density Matrix . . . 49
10.2 A more complex case: the hexagonal Argon crystal . . . 50
10.2.1 The recovery of the information from PROPERTIES . . . 50
10.2.2 The excitations domains . . . 54
10.2.3 The Dual Basis Set option . . . 56

11 Theory in pills **57**
11.1 The local approach . . . 57
11.1.1 Local description of the occupied manifold: Wannier Functions . . . 58
11.1.2 Local description of the virtual manifold: Projected Atomic Orbitals . . . 60
11.1.3 Symmetry properties of WFs and PAOs . . . 61
11.1.4 Wannier functions: distances and domains . . . 63

11.2 Local-MP2 method for periodic systems 66
11.2.1 Local MP2 Energy 66
11.2.2 Local MP2 amplitudes 67
11.2.3 Lennard-Jones extrapolation for very distant pairs 69
11.2.4 The $(\mathbf{ia}|\mathbf{jb})$ integrals 71
11.3 LMP2 correction to the Hartree-Fock Density Matrix 76

Bibliography **78**

Subject and Keywords Index **80**

Chapter 1

Generalities

1.1 Introduction

CRYSCOR is a **post-Hartree-Fock** (HF) **local**-correlation program for crystalline solids [1, 2]. Since the reference HF solution is provided by the CRYSTAL code [3], it can be considered as the post-HF option of it.

The post-HF method currently implemented is a perturbative method, namely Møller-Plesset at the second-order (**MP2**). Well localized symmetry adapted Wannier Functions (WF) [4, 5, 6] are adopted instead of delocalized Bloch functions for the description of the occupied manifold; this permits the exploitation of the **short-range** nature ($E \propto r^{-6}$) of electron correlation following the general Pulay scheme [7, 8] as implemented in the molecular MOLPRO code [9]. Its generalization to periodic 3D systems requires the full exploitation of both point and translational symmetry and permits $\mathcal{O}(\mathcal{N})$ scaling of computational costs, $\mathcal{N}$ being the size of the unit cell of the crystal. The essentials of this method are described in Chapter 11.

Note that the localization procedure and the MP2 technique implemented in CRYSCOR are not suitable for open-shell systems, for conductors, or for semiconductors with very small gap.

1.2 Conventions

In the description of the input data the following notation is adopted:

-	•	new record
-	[ ]	default values.
-	{ }	suggested values.

The A, I and F letters label alphabetic, integers and float data type, respectively.

All keywords are entered with a case insensitive format; they must be typed left-justified with no leading blanks.

The following acronyms are used (see Chapter 11 for details):

-	**AO**	Atom-centred Atomic Orbital obtained as a linear combination of Gaussian Type Orbitals (GTO)
-	**WF**	Wannier Functions, basis set for the occupied manifold
-	**WFf**	"Flower" of symmetry adapted WFs, basis for a representation of the point group
-	**PAO**	Projected Atomic Orbitals, basis set for the virtual space
-	**FF**	Fitting Functions for the Density Fitting approximation of two-electron integrals

1.3 Functionality

The basic functionality of the code is outlined below.

Local features

- Hierarchical treatment of biexcitations
- Definition of the virtual manifold according to different criteria
- Lennard-Jones extrapolation for evaluating long-range energy contributions

Evaluation of bielectronic integrals

- Exact method
- Periodic Density Fitting technique in both direct and reciprocal space
- Multipolar expansion

One-electron Density Matrix

- MP2 correction to the HF Density Matrix feasible
  (the MP2-corrected Density Matrix can be passed to the PROPERTIES code)

Basis set

- Automatic tranfer of Basis set information from CRYSTAL to CRYSCOR
- Dual Basis set option for improving the description of the virtual manifold

1.4 Citing Cryscor

When a user presents to the scientific community (publications or congresses) any result obtained with CRYSCOR, he has to cite the program in the proper way as follows:

- C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz and D. Usvyat, *J. Comput. Chem.*, **29**, 2113 (2008); A. Erba and M. Halo, *CRYSCOR09 User's manual*, University of Torino, Torino, 2009 (www.cryscor.unito.it).

1.5 The Molpro contribution

The Density Fitting module of the CRYSCOR program is based, upon license, on the corresponding module of the MOLPRO molecular program [F. R. Manby, P. J. Knowles, and A. W. Lloyd, *J. Chem. Phys.* **115**, 9144 (2001)].

1.6 How to run a CRYSCOR job

To execute a CRYSCOR job, you have to run two preliminary calculations. The first one is a CRYSTAL calculation, performed at HF level; from this run, along with the output file (*filename1.out*), you get the unformatted external unit 9 (*fort.9*) or the formatted external unit 98 (*fort.98*) which contain information about the geometry, symmetry, computed wavefunction, Fock and Density matrices. The use of the **TOLMP2** keyword, which sets the tolerances of the HF calculation to convenient values for the subsequent correlation calculations, is recommended. As a second step you have to run a PROPERTIES calculation for transforming the canonical Crystalline Orbitals into localized Wannier functions; in order to do that, you need feeding the *fort.9* (or *fort.98*) unit to PROPERTIES. This second run generates the unformatted external unit 80 (*fort.80*). A CRYSCOR run needs both units 9 and 80.

The formatted unit 98 can be transformed into the corresponding unformatted unit 9 by means of the **RDFMWF** keyword of the PROPERTIES program.

For more details about how to run a CRYSTAL calculation please refer to *Crystal User's Manual* [3] and to CRYSTAL web page: www.crystal.unito.it.

On Unix systems, the CRYSTAL, PROPERTIES and CRYSCOR executables can be called via the standard scripts: **runcry09**, **runprop09** and **runcryscor09**. A common sequence of commands is schematized in what follows:

A. runcry09 *filename1*
The input file *filename1.d12* must be there; this script launches the CRYSTAL calculation and automatically saves in your working directory both the output file *filename1.out* and the unit 9 as *filename1.f9*.

B. runprop09 *filename2 filename1*
Both the PROPERTIES input file (*filename2.d3*, see below, Section 2) and unit 9 (*filename1.f9*) must be there. This script saves the output file (*filename2.outp*) and unit 80 as *filename2.f80*; this unit contains information about the localized and symmetrized WFs.

Program	input file	output file	unit saved
CRYSTAL	.d12	.out	.f9, [.f98]
			[.f78 (see page 31)]
PROPERTIES	.d3	.outp	.f80
			[.f81 (for WFs restart, see CRYSTAL Manual)]
CRYSCOR	.d4	.outc	[.f84 and .f151 (see page 25)]
			[.f131, .f132 and .f133 (see page 24)]
			[.f63 (see page 32)]

Table 1.1: A summary of the extensions of input/output files and external units.

C. runcryscor09 *filename3 filename1 filename2*
Each CRYSCOR run needs an input file (*filename3.d4*), and the external units 9 and 80 (*filename1.f9* and *filename2.f80*). This script generates the output file *filename3.outc* and some external units that can be useful for later calculations (see Table 1.1 for more information).

Once the **runcry09** and **runprop09** scripts have been correctly set up (i.e. when the necessary environment variables have been defined), you can use the **runcryscor09** script without setting any new variable; the only thing you should take care of is to put your `cryscor` executable in the same directory of your `crystal` and `properties` executables.

1.7 Bugs Reporting

The authors would greatly appreciate comments, suggestions and criticisms by the users of CRYSCOR; consult our forum on the CRYSCOR website www.cryscor.unito.it for more details. When reporting a bug, the user is asked to provide all the input and output files (*.d*12, *.out*, *.d*3, *.outp*, *.d*4 and *.outc*) along with the external fortran units 9 and 80.

1.8 Acknowledgments

The authors are grateful to Prof. Claudio Zicovich-Wilson and to Prof. Roberto Dovesi.

Chapter 2

The main ingredient: Wannier Functions

The post-HF correlation method which is currently implemented in the CRYSCOR program is the MP2 one reformulated according to the so-called *local approach* which implies the use of localized functions in the description of the occupied and virtual manifolds. The accurate description of the occupied space has of course a great relevance in the determination of the correlation energy.

In order to describe the occupied manifold, Wannier Functions (WF) are used instead of delocalized Bloch Functions (BF). While BFs are translationally invariant apart from a phase factor, WFs are translationally equivalent. Some further information about WFs is provided in Chapter 11.

A PROPERTIES input file (with extension *.d3*, see Table 1.1) has to be prepared in order to obtain well localized and Symmetry Adapted WFs (SAWF). The definition of WFs is controlled by an input block, opened by the **LOCALWF** keyword and containing a set of keywords, as explained in the *Crystal User's Manual* [3]; a mandatory keyword among them is **VALENCE**, since we only perform valence-electron correlation.

A typical input file for the localization/symmetrization procedure looks like that:

`NEWK`	Diagonalization of the Fock matrix
`8 8 8`	Shrinking factor
`1 0`	Fermi energy is computed
`LOCALWF`	Start of the localization input block
`VALENCE`	Only valence orbitals are localized
`SYMMWF`	Start of the symmetrization input block
`END`	End of the symmetrization input block
`END`	End of the localization input block
`END`	End of the properties input block

Table 2.1: The simplest possible localization input.

The *a posteriori* symmetrization procedure, mandatory in the case of a subsequent MP2 calculation with CRYSCOR, is activated by means of the **SYMMWF** keyword. This procedure [6] and the corresponding symmetry-based classification of WFs are discussed in the following section.

2.1 The symmetrization step

The symmetrization of the WFs is a computationally convenient step since it allows the full exploitation of the point-symmetry of the system and at the same time it reduces the numerical noise of the computed values.

For the time being, the WFs are symmetrized after the localization step that is, by following a so-called *a posteriori* scheme which can be outlined as follows:

1. after the localization step, the WFs are centered at different sites of the reference cell which are invariant with respect to point-symmetry subgroups H of the space group G of the crystal;

2. for each site, a coset decomposition of G can be performed, induced by H, thus leading to the definition of $N_F = \frac{|G|}{|H|}$ symmetry operators (coset representatives) which rotate the site into equivalent ones;

3. among each set of equivalent sites, a reference site is chosen;

4. the WFs of each reference site are symmetrized according to the corresponding invariant subgroup H: such WFs turn out to be basis functions of the irreducible representations (IRREP) of the subgroup H and hereafter will be referred to as *petals*;

5. the collection of *petals* which constitute a basis for a given IRREP of H constitutes a so-called *flower*. It is worth noting that bi/tri-dimensional IRREPs give rise to *flowers* made up of two/three *petals*, respectively;

6. the rotation of the reference *flower* performed by means of the corresponding N_F-1 coset representatives (identity excluded) yields the creation of other N_F-1 symmetry related *flowers*; the set of such equivalent *flowers* constitutes a *bunch*;

7. in general, more than one *bunch* could be associated to the same reference site;

8. all the SAWFs of the reference cell are then replicated in the other cells of the crystal;

9. as a result of this procedure, each WF is fully classified by four indices $|\mathtt{b,f,p,g}\rangle$ ($\mathtt{b}$ = bunch, $\mathtt{f}$=flower, $\mathtt{p}$= petal, $\mathtt{g}$= crystal cell) such that a general symmetry operator of the system $\widehat{R} \in$ G, applied to a WF gives: $\widehat{R}\ |\mathtt{b,f,p,g}\rangle = \sum_{p'} [A^b(R)]_{pp'}\ |\mathtt{b,f}^R\mathtt{,p',g}^R\rangle$.

2.1.1 An example: the three-layers slab of MgO

In order to help the user in the comprehension of the symmetry relations which occur among the symmetrized WFs, we present in some detail a simple example: a three-layers slab of MgO. The structure of this system is sketched in Figures 2.1 and 2.2 along with the schematic representation of the symmetry properties of the valence WFs of the system.

The unit cell of the system contains two atoms per layer: one Oxygen (O) and one Magnesium (Mg). Since we are neglecting core WFs, the Oxygen atom only hosts the valence WFs of the

system: four per O atom, one s-type, one p_z-type, one p_x-type and one p_y-type. The total number of electrons per cell is 24 and the number of WFs per cell is 12.

The 12 WFs are centered at three distinct sites: the O atom of the top layer (O1), the O atom of the central layer (O2) and the O atom of the bottom layer (O3). The O2 atom belongs to a site which is characterized by the whole point-symmetry of the system (D_{4h}). According to what we have stated before, each *flower* centered on it has to be invariant under the action of any symmetry operator of the Point Symmetry Group P of the crystal; as a consequence, this site can not be symmetry related to any other one in the crystal. The O1 and O3 atoms belong to two symmetry-related sites characterized by an invariant subgroup of symmetry H (C_{4v}) whose order is half the order of P; it follows that each *flower* centered on O1 can be rotated into an equivalent one centered on O3 by means of a symmetry operator (the representative of the coset decomposition of G, induced by H, that is the mirror plane which coincides with the central layer of the slab).

The O1 and O2 sites are then taken as the reference ones. Please, refer to Figure 2.1 for a graphical legend of the sites.

The symmetry classification of the 12 WFs is the following, as printed at the beginning of a CRYSCOR calculation:

BUNCH	FLOWER ABS	FLOWER REL	PETAL	TYPE	ATOM1	(G1)	-	ATOM2	(G2)
1	1	1	1	NON BONDING	O 1	(1)	-		
1	2	2	2	NON BONDING	O 3	(1)	-		
2	3	1	5	NON BONDING	O 1	(1)	-		
2	3	1	6	NON BONDING	O 1	(1)	-		
2	4	2	4	NON BONDING	O 3	(1)	-		
2	4	2	8	NON BONDING	O 3	(1)	-		
3	5	1	7	NON BONDING	O 1	(1)	-		
3	6	2	9	NON BONDING	O 3	(1)	-		
4	7	1	3	NON BONDING	O 2	(1)	-		
5	8	1	10	NON BONDING	O 2	(1)	-		
5	8	1	11	NON BONDING	O 2	(1)	-		
6	9	1	12	NON BONDING	O 2	(1)	-		

From the analysis of the atomic orbital populations of the 12 WFs (hereafter referred to as *petals*), as reported in the output of the PROPERTIES calculation, we can recognize the character of each

petal: 1 3 2 are the s-type orbitals, 7 12 9 are the p_z-type orbitals, 5 4 10 are the p_x-type orbitals and 6 8 11 are p_y-type orbitals of O atoms 1 2 3, respectively.

The symmetry of the system gives rise to 6 *bunches* centered on three different sites (the three O atoms) for a total of 9 *flowers*.

There are three *bunches* centered on O2 (4, 5 and 6), each one made up of one single *flower*: two of these *flowers* (7 and 9) just have one *petal* (the s-type and p_z-type orbitals, respectively) while the *flower* number 8 has two *petals* (the p_x-type and p_y-type orbitals). This means that the four *petals* of O2 have been symmetrized according to two one-dimensional (A and B) and one bi-dimensional (E) IRREPs of P.

Figure 2.1: Side view. Schematic representation of the symmetry properties of the SAWFs of a three-layers slab of MgO. Small red and grey circles represent Oxygen and Magnesium atoms, respectively. Blue circles, red arrows and green ovals represent s-type, p_z-type and $p_{x,y}$-type SAWFs, respectively. *Petals* are labelled according to the numeration reported in the Table above.

Let us comment on the symmetry classification of the *petals* centered on the O1 and O3 sites. Symmetry-related *flowers* constitute a *bunch*; there are three *bunches*, of two *flowers* each, which involve these two sites. Two *bunches* (1 and 3) are made up of two *flowers* of one *petal* (the two s-type and p_z-type orbitals of O1 and O3), while the last *bunch* (2) is made up of two *flowers* of two *petals* (the two p_x-type and p_y-type orbitals of O1 and O3).

A schematic representation of the symmetry properties of such orbitals is reported in Figures 2.1 and 2.2. The two-*petal* structure of the *flower* 3 is clarified from Figure 2.2. When acting over one of these petals ($|p_1\rangle$ or $|p_2\rangle$) with a symmetry operator $\widehat{R}$ of the invariant subgroup H, one gets:

$$\widehat{R}|p_j\rangle = \sum_{i=1}^{2}[A^E(R)]_{ji}|p_i\rangle$$

Figure 2.2: Top view. Schematic representation of the symmetry properties of the SAWFs of a three-layers slab of MgO. The top layer (layer 1 in Figure 2.1) is reported which contains the site O1. Conventions as in Figure 2.1

where E labels the used bi-dimensional IRREP of the subgroup and $A^E(R)$ are the corresponding unitary matrices.

Chapter 3

Input format and mandatory keywords

3.1 Input format

The CRYSCOR input consists of a series of keywords (KW) possibly followed by the respective arguments, to be written in separate cards in free format. The following sections will describe these KWs in detail along with their corresponding arguments (the kind of data to be inserted is specified by three letters: *I* for integers, *F* for real-float and *A* for alphabetic).

Not all KWs are necessary; the ones that are mandatory are marked as **KM**, the others (i.e. the optional ones) as **KO**. The ones marked with **KT** are technical keywords, used by the developers of the code and their use is not recommended to the user. For the arguments introduced by optional keywords, default values are provided by CRYSCOR, as indicated within square brackets.

The order in which the KWs have to be inserted in the input file is practically free, but some conditions must be respected:

1. **KNET** must be the first keyword[1];
2. **MEMORY** must be the second keyword, specifying the maximum available memory available on the machine running the program (in Mbytes);
3. One of the KWs aimed at the definition of excitation domains has to be inserted (**DOMPUL**, **DOMMOL**, **DOMDEF**);
4. The use of the **DFITTING** keyword which activates the Density Fitting approximation for the computation of the bi-electronic integrals is *strongly recommended*, although not mandatory since in principle one could solve these integrals exactly.
5. **END** must be the last keyword of the input.

The description of the input is subdivided into sections which control different parts of the code. Inside each section, the meaning of KWs and of the respective parameters is summarily indicated in an input deck; then, detailed explanations, suggestions and comments are added and some input examples are provided.

[1] except when **DUALBAS** option is active, see Section 6

The simplest possible CRYSCOR input file is reported below (Table 3.1). In order to understand which of the many available KWs are of effective interest in ordinary applications, we report in Table 3.2 an input equivalent to the previous one, but with the explicit definition of some important parameters, with their default values.

`KNET`	Construction of the HF Density Matrix
`12`	Shrinking factor
`MEMORY`	Memory required
`2500`	Value in Mbytes
`DOMPUL`	Definition of the excitation domains
`0.98`	T^{BP} parameter (see Sec. 11.1.4)
`DFITTING`	Density Fitting input block
`DIRECT`	Direct-space technique
`PG-VTZ`	DF auxiliary basis set
`ENDDF`	End Density Fitting input block
`END`	End of CRYSCOR input

Table 3.1: The simplest possible CRYSCOR input. In this way many default values are used.

`KNET`	Construction of the HF Density Matrix
`12`	Shrinking factor
`MEMORY`	Memory required
`2500`	Value in Mbytes
`DOMPUL`	Definition of the excitation domains
`0.98`	T^{BP} parameter (see Sec. 11.1.4)
`PAIR`	Definition of the hierarchy of WF-pairs
`8.  12.`	d_1 and d_2 as defined in Sec. 11.1.4
`DFITTING`	Density Fitting input block
`DIRECT`	Direct-space technique
`PG-VTZ`	DF auxiliary basis set
`ENDDF`	End Density Fitting input block
`MULTIPO`	Multipolar approximation activated
`4`	Multipoles up to hexadecapoles are used
`TOBJ`	Truncation of WF and PAOs as LCAO
`0.0001 0.0001`	See Sec. 11.1.1
`LENJONES`	A Lennard-Jones (LJ) extrapolation is performed
`6.`	LJ parameter (See Sec. 4.1.3)
`END`	End of CRYSCOR input

Table 3.2: This is a CRYSCOR input fully equivalent to the previous one; the only difference is that some default options have been explicitly written.

3.2 General and mandatory keywords

The **KNET** keyword creates a new mesh of points in the reciprocal space where the HF Density Matrix is evaluated; this matrix will be used in the construction of the PAOs. The suggested value of the new shrinking factor is the same as in the CRYSTAL calculation.

Via the **MEMORY** keyword the user has to declare how much memory the CRYSCOR executable can use. If a low memory value is given, CRYSCOR will try to reduce its need for memory by increasing the use of disk; a reasonable value is a bit less than the amount of available memory of the machine the user is working with.

The CRYSCOR input file must be closed with the **END** keyword.

rec		variable	value	meaning
•		**KNET**	**KM**	The HF Density Matrix is generated by sampling the reciprocal space at a number of **k**-points
•	I	IS		Shrinking factor for reciprocal space net (Monkhorst net)
•		**MEMORY**	**KM**	Memory allocation
•	I	IMEM		Allocation in MBytes
...				
•		**END**	**KM**	To end the main input stream

Chapter 4

Occupied and virtual space description

4.1 Occupied space description

The nature of Wannier functions is described in detail in Chapter 2 as well as in Chapter 11. What we want to point out here is the importance of the concept of pairs of WFs. As a matter of fact, the total LMP2 energy per cell can be written as a sum over LMP2 energies per WF-pair: $E^{LMP2} = \sum_{i_0 j} E^{LMP2}_{i_0 j}$, where i,j run over WFs. Given that the first WF in the pair belongs to the 0-cell (0 pedix of i_0 in the formula), it becomes apparent how crucial is the definition of the distance of the second WF, i.e. the distance up to which two electrons "see" each other. The definition of a distance criterion for WF pairs allows furthermore a hierarchical treatment of the pairs: a series of approximated techniques for the computation of the bielectronic integrals involved is possible, as shown in section 5.1.

In this section the KWs controlling the fundamental parameter of WF-WF pairs distance are reported.

4.1.1 Definition of kernels

As we discuss in some detail in Section 11.1.4, a *kernel* is associated to each WF (i) which consists in a set of atoms.

Kernels serve only to specify the distance D_{ij} between two WFs, defined as the *minimum distance between any two atoms belonging to the respective kernels.*

Kernels are defined according to a unique scheme related to a Mulliken population (MP) analysis; only those atoms are included in the kernel of a given WF whose contribution in terms of MP to that WF is greater than a given threshold Q^0 (see equation 11.17). This threshold is usually set to 0.1 but it can be modified via the **MINPOP** keyword (not advisable!). According to the *atomic* or *bond* character of a given WF, the kernel usually contains one single atom or the two atoms involved in the bond, respectively. The default values are strongly recommended.

rec	variable	value	meaning
•	**MINPOP**	**KT**	Designs *kernels*, used in the classification of WF-pairs
• F	MINPOP	[0.1]	Only those atoms are included in the *kernel* whose Mulliken population is > `MINPOP` (see equation 11.17)

4.1.2 Pair selection

The **PAIR** keyword introduces a hierarchical treatment of (i, j) WF pairs according to their distance D_{ij} (defined as above as well as in Section 11.1.4) by means of the two parameters d_1 and d_2 specified in Å in the second card. The first is always in the zero cell. Only WF-pairs for which i belongs to the reference zero cell and $D_{ij} \leq d_2$ are retained in the calculation. The considered ones are further classified into *strong* ($D_{ij} = 0$) *weak* ($0 < D_{ij} \leq d_1$) and *distant* ($d_1 < D_{ij} \leq d_2$) pairs. For most purposes, strong and weak pairs are treated in CRYSCOR as a unique set, that of *close-by* pairs. Different approximated techniques for the computation of the bielectronic integrals involved is possible, see section 5.1.

rec	variable	value	meaning
•	**PAIR**	**KO**	Partition of the occupied space
• 2F	d_1 d_2	[8 12]	Values in Å

4.1.3 Lennard-Jones extrapolation

One of the distinctive features of local correlation schemes in molecular calculations is that one can safely ignore the contributions E_{ij} from pairs which are "very distant" from each other, that is, when the distance between the respective core domains exceeds a pre-fixed value d_2 (typically, $d_2 = 12$ Å). This approximation is crucial to warrant N-scaling of computational costs, and is justified by the fact that at large distances pair energies follow the London d_{ij}^{-6} law as is present in the well known 6-12 Lennard-Jones expression, d_{ij} being the distance between the centers of the two distributions. However, in 3D crystals the number of pairs grows quadratically with the distance d and the contribution to the correlation energy from all pairs at a distance $d_{ij} \geq d$ falls off merely as d^{-3}. Considerably larger cutoff distances d_2 should be adopted with respect to molecular calculations in order to achieve comparable accuracy, with a formidable impact on computational costs. This problem is easily circumvented by taking advantage of the simple law of inter-pair decay rate and of translational symmetry so as to extrapolate to infinity the sum of far-off contributions. The algorithm implemented for this purpose is described in detail on page 69.

rec	variable	value	meaning
•	**LENJONES**	**KO**	Lennard-Jones extrapolation to infinity
• F	d_{LJ}	[6]	Only WFf-pairs with distance from the reference cell comprised between d_{LJ} and d_2 (see above) are used to calculate Lennard-Jones coefficients

4.1.4 Pair partitioning in adsorption problems

A case of widespread interest is that where a molecule is interacting with an adsorbing structure, for instance a 2-d substrate or a porous 3-d system. CRYSCOR allows a special classification of ij pairs in such cases, according to the subystems i and j belong to. The related KWs are as follows. The **MOLATOMS** KW defines the molecular fragment. Molecular (*mol*) WFs are then identified as those which have at least one *molecular* atom in their excitation domain; all others are environmental (*env*) WFs. WF-pairs can then be of three types: *mol-mol*, *env-env*, *mol-env*. The **MOLPAIR**, **ENVPAIR**, **MOENPAIR** KWs allow different values of d_1 and d_2 to be assigned to these three types, respectively.

rec	variable	value	meaning
•	**MOLATOMS**	**KT**	Defines the partition into "molecule" and "environment"
• I	n	[0]	Number of atoms of the "molecule"
• nI	i=1,n		List of the n atoms belonging to the "molecule" (their labelling is the same as in the CRYSTAL input).
•	**MOLPAIR**	**KT**	Strong/weak/distant WF-pairs structure for *mol-mol* pairs
• 2F	d_1 d_2		d_1 and d_2 are expressed in Å. If d_1 and d_2 are 0, no pairs of this type are included.
•	**ENVPAIR**	**KT**	Strong/weak/distant WF-pairs structure for *env-env* pairs
• 2F	d_1 d_2		d_1 and d_2 are expressed in Å. If d_1 and d_2 are 0, no pairs of this type are included.
•	**MOENPAIR**	**KT**	Strong/weak/distant WF-pairs structure for *mol-env* pairs
• 2F	d_1 d_2		d_1 and d_2 are expressed in Å. If d_1 and d_2 are 0, no pairs of this type are included.

Note: These KWs can be useful, but great attention has to be paid when using them. For instance, problems can arise when the molecule is too close to the surface.

A detailed description of these KWs along with some general remarks about the modeling of adsorption phenomena are reported in the guide *Adsorption of molecules on surfaces with* CRYSCOR09 by D. Usvyat that can be found in the CRYSCOR09 website (http://www.cryscor.unito.it) at the *Supporting Material* section [10].
Also note that there is a difference in the syntax of these KWs with respect to **PAIR**. With **PAIR**, setting to 0.0 both d_1 and d_2 implies that only *Strong* WF-pairs are computed, whereas here it means that no WF-pairs are considered of the selected type.

4.2 Virtual space description

Once set up the description of the occupied manifold, the virtual manifold needs to be dealt with. The localized functions which span the virtual space are the *Projected Atomic Orbitals* (PAO), each obtained by projecting out of a given AO the portion that belongs to the occupied space (see Section 11.1.2).

PAOs are labelled with the letters a,b,...; each of them is associated to (or belongs to) the atom where the corresponding AO is centered.

4.2.1 (Excitation) Domains: General definitions

For each pair of WF, only those bi-excitations $[ij] \rightarrow [ab]$ are considered in CRYSCOR which consist in promoting the two electrons from the i,j WFs to two spatially close PAOs a,b. For this purpose, we associate to each WF i (or more precisely, to each *flower* of WFs, see below) an *excitation domain* D_i, which consists in a **set of atoms**. For a bi-excitations $[ij] \rightarrow [ab]$ to be considered, **both a and b must belong to one of the atoms of the union of the two domains, $\mathbf{D}_i \cup \mathbf{D}_j$**.

The definition of the excitation domains is crucial to the accuracy and cost of the calculation. For this reason, *no default is automatically provided by the program* and the user has to explicitly define shape and size of the excitation domains (hereafter simply "domains").
Several distinct schemes can be used for this purpose, which will be introduced and commented on in the next sections:
- Mulliken population analysis of the WFs (**DOMPUL**), see Section 4.2.2;
- Atomic distances from the WF-centroids (**DOMDEF** or **DOMDEF2**), see Section 4.2.3;
- Molecular domains (**DOMMOL**) for molecular crystals only, see Section 4.2.4.

One of these KWs is mandatory, so they are marked **KO(M)**.
Alternatively, one can use a definition based on the coefficients in the AO expansion of the WF (**DOMCOE**, see Section 4.2.5) but the corresponding KW is classified as **KT** because its use is more delicate.
Let us underline here that WFs belonging to the same "flower" (that is, which transform into each other under action of the invariant site symmetry subgroup) have the same domain. Therefore, in a sense we have to define flower domains. Furthermore, each flower domain is symmetrized according to the corresponding site subgroup.

A general KW (**MAXDOM**) allows the user to restrict anyhow the number of atoms in any excitation domain to a maximum value (MAXD).

rec	variable	value	meaning
•	**MAXDOM**	**KO**	Set the maximum number of atoms that can enter the definition of an excitation domain
• F	MAXD	[60]	Maximum number of atoms in the domains

4.2.2 Domain definition: Mulliken Population Analysis

A definition which is chemically sensible is provided by the Boughton-Pulay (BP) criterion presented in some detail in Section 11.1.4: a sort of "Mulliken population analysis" is performed for each WF, by looking how much of it (as a fraction of 1) is "provided" by the different atoms of the system; these are included in the domain in order of decreasing contribution, until the domain contains a preset fraction (T^{BP}) of the total WF-population. Note that atoms which are symmetry equivalent with respect to the flower symmetry subgroup, and therefore contribute with the same amount of population to the WF, are all included in the domain (or all excluded).

rec	variable	value	meaning
•	**DOMPUL**	**KO(M)**	Boughton-Pulay (BP) domains
• F	T^{BP}	[0.98]	Total WF-population in BP domains (see Section 11.1.4)

4.2.3 Domain definition: Stars of neighbors

A more specific tailoring of flower-domains is possible by activating the following keywords.

With the **DOMDEF** keyword, for each flower, stars of symmetry-equivalent atoms are included in the domain; the total number of stars (`NS`) and the "identification labels" (`ISTAR`) of the different stars included (in order of increasing distance from the center of the flower) have to be inserted as arguments. Note that the distances of the neighboring atoms are defined according to the flower centroid that is, the center of mass of the centroids of all the WFs belonging to the flower.

rec	variable	value	meaning
	insert one of the following set of cards, I or II:		I
•	**DOMDEF**	**KO(M)**	Designs flower-domains by **stars of atoms**
• I	NF		Number of flowers
	for each flower F=1,NF insert the following values:		
• 2I	F NS		Selected flower; Number of stars of atoms to be included
• *NS*I	ISTAR=1,NS		Labels of the `NS` stars of atoms (according to a distance criterion with respect to the center of the flower)
	or		II
•	**DOMDEF2**	**KO(M)**	Designs flower-domains by the **number of atoms**
• I	NF		Number of flowers
• *NF*I	NAT=1,NF		Number of atoms to be included in the `NF` domains

When using the **DOMDEF** definition, be careful that the domains for flowers of different bunches are really consistent; depending on the system and on the localization procedure, it may happen that when a WF is not exactly centred on the atom/bond the star of neighbors is splitted and therefore a further star(s) need to be declared in the input. See section 10.2 for an example. When the flowers that result with a different domain size belong to the same bunch, instead, CRYSCOR automatically levels the size of the domain to the largest one.

As an alternative, by means of the **DOMDEF2** keyword, the user has to indicate how many atoms (`NAT`) have to be included into each flower-domain. The program will automatically select for each flower the `NAT` atoms closest to the center of the flower. If needed, additional atoms are automatically added in order to fill the outermost star of atoms. See section 10.2 for an example.

Some information about the flowers (their number, NF, their reference number, their main features) is printed by the PROPERTIES code at the end of the localization procedure, and also at the beginning of the CRYSCOR output (look at the string `WF INFORMATION`).

4.2.4 Domain definition: Molecular domains

When we are dealing with molecular crystals (with m molecules M_1, M_2, ..., M_m in the reference cell) it can be convenient to work with molecular domains that is to say, to include in each domain of a WF belonging to the *ith*-molecule M_i the whole set of atoms of that molecule.

No arguments have to be provided.

rec	variable	value	meaning
•	**DOMMOL**	**KO(M)**	Molecular domains

4.2.5 Domain definition: Coefficients of WFs

The domain of the WF $\phi_i = \sum_\mu c_{\mu,i}\chi_\mu$ is specified as comprising all atoms which contain at least one AO χ_μ such that the corresponding coefficient $c_{\mu,i}$ is larger in absolute value than a prescribed value: by default this threshold (`TDOM`) is set to 0.01. One of the main disadvantages of this method is the fact that the value of `TDOM` heavily affects the size of the domain, which can easily become too large. We suggest not to use this KW, except for tests.

rec	variable	value	meaning
•	**DOMCOE**	**KT**	Domains by atomic coefficients
• F	TDOM	[0.01]	Those atoms A are included in the domain of the general WF (i) for which $\lvert c_{\mu,i}\rvert >$TDOM for at least one $\mu \in A$

4.2.6 Checking the shape of domains

It may sometimes be useful to have a preliminary look at the shape of the domains without performing a full calculation. This possibility is provided by the **TESTDOM** keyword.
The coordinates and types of the atoms included in each domain are stored on Fortran unit 17 in `xyz` format for visualization, and the computation stops after the domain calculation. Moreover, the composition of the different stars, their distance from the center of the flowers, etc., are printed in the output file thus allowing further selection (look at the string `DOMAIN INFORMATION`).

Two choices are possible. Either the same number and sequence of stars are used for all the

rec	variable	value	meaning
•	**TESTDOM**	**KO**	Test run for checking flower-domains
	insert one of the following set of cards, I or II:		I
•	**ALLEQ**		All flower-domains are equal
• I	NF		Number of flowers
• I	NS		Number of stars
• *NS*I	ISTAR=1,NS		Labels of the `NS` stars of atoms (according to a distance criterium with respect to the center of the flower)
	or		II
•	**NONEQ**		Different flower-domains
• I	NF		Number of flowers
	for each flower F=1,NF insert the following values:		
• 2I	F NS		Selected flower; Number of stars of atoms to be included
• *NS*I	ISTAR=1,NS		Labels of the `NS` stars of atoms (according to a distance criterion with respect to the center of the flower)

domains (keyword **ALLEQ**), or differences are allowed between the different domains (keyword **NONEQ**). In the former case, only one variable `NS` (number of stars) and one vector `ISTAR` (labels of the selected stars) must be defined. In the latter case, domains are generated following the same criterion as for **DOMDEF**.

4.3 Truncation criteria

The **TOBJ** keyword allows truncating the tails of WFs and PAOs in their AO expansion: all coefficients smaller in an absolute value than `TWF` and `TPAO` in the respective expansion are set to zero and therefore disregarded.
LONTOL is a threshold for excluding linearly dependent virtual functions (PAOs - Projected Atomic Orbitals) from the LON (orthonormalized PAOs) set of each given WF-pair (see Section 11.1.2). Note that too small values of **LONTOL** can result in catastrophic behaviour.
Two technical keywords allowing more specific truncation criteria to be assigned are **COREQ** and **DELPAO**.

rec	variable	value	meaning
•	**TOBJ**	**KO**	Truncation of WFs and PAOs
• 2F	TWF, TPAO	[0.0001 0.0001]	Threshold of the coefficients
•	**COREQ**	**KT**	Exclude "evanescent" PAOs
• I	JNORM	[1]	Eliminate PAOs with norm $< 10^{-\mathrm{JNORM}}$
•	**LONTOL**	**KT**	Truncation of LON
• F	t_σ	[0.0001]	Tolerance on the LON overlap matrix
•	**DELPAO**	**KT**	Exclude the selected PAOs from the virtual space
• I	N		Number of PAOs to be eliminated
• *M*I	P=1,N		PAOs to be eliminated

4.4 Fixing of indices for manual geometry optimization

A procedure for the automatic geometry optimization at LMP2 level is not currently available in the CRYSCOR code, although the implementation of the analytical gradients of the LMP2 energy is under development and will appear in the next public version of the program.

For the time being, only discrete (i.e. by hand) geometry optimizations are possible i.e. the user has to define by his own a discrete set of nuclear configurations (let us say N_{conf} configurations) which have to be investigated separately by means of N_{conf} single point calculations. As a matter of fact one of the most delicate steps of the procedure is exactly the construction of the N_{conf} nuclear configurations. However, at least in some cases, this step turns out to be not so complicated: this is the case, for instance of lattice parameters and adsorbate/surface distance optimizations.

As usual, when such a discrete optimization is performed, one has to freeze some quantities that have to be the same in all points in order to avoid any inconsistency among them which may lead to a not very smooth optimization curve. For instance, the **FIXINDEX** keyword has to be used in CRYSTAL calculations and the **FIXWF** keyword has to be used in the PROPERTIES calculations. See the CRYSTAL User's Manual for details [3].

Analogously, some quantities have to be fixed also in the CRYSCOR calculations. In particular, one has to freeze the WF-pairs classification according to distance (**FIXPAIR** keyword) and the size and shape of the excitation domains (**FIXDOM** keyword) . These two keywords have to be used in the calculation performed on the **most dense geometry** (i.e. with the smallest lattice parameter or adsorbate/surface distance, for instance) and generate two corresponding external units: 133 and 131, respectively.

For all the other $N_{\text{conf}} - 1$ single point calculations, the **PAIRREA** and **DOMREA** keywords have to be inserted in the CRYSCOR input and the two units 133 and 131 obtained with the reference geometry have to be provided. **Note:** in some cases, it can be also convenient to fix the number of PAOs by means of the **FIXNPAOS** and **NPAOSREA** technical keywords.

rec	variable	value	meaning
•	**FIXPAIR**	**KO**	Saving the information on the number and type of WF-pairs in the external unit 133
•	**FIXDOM**	**KO**	Saving the information on the domains in the external unit 131
•	**FIXNPAOS**	**KT**	Saving the number of PAOs in the external unit 132
•	**PAIRREA**	**KO**	Reading the information on the number and type of the WF-pairs from the external unit 133
•	**DOMREA**	**KO**	Reading the information on the domains from the external unit 131
•	**NPAOSREA**	**KT**	Reading the number of PAOs from the external unit 132
•	**DOMONLY**	**KO**	Does not perform the actual LMP2 calculation but obtains the information on pairs, domains and PAOs and stops

Chapter 5

LMP2 integrals and equations

5.1 Two-electron integrals

In this section we discuss one of the most essential issues of the LMP2 method, the evaluation of the Electron Repulsion Integrals (ERIs) (**ia** | **jb**), which enter the definition of both the energy and amplitude (Section 11.2.2) equations. The techniques employed in CRYSCOR for dealing with this problem differ according to the distance D_{ij} between the two WFs, **i** and **j**.

For *close-by* pairs ($D_{ij} \leq d_1$) the *exact*, i.e. analytic, evaluation of the ERIs is the *default option*. Some technical KWs may be specified, as explained below. Note however that this technique is as a rule terribly time-consuming and should only be considered for calibration purposes.

The **strongly recommended** option for *close-by* pairs is a Density Fitting Periodic technique (DFP), as described in Section 11.2.4. The corresponding input is commented on in details below (Section 5.1.3). For *distant* pairs ($d_1 < D_{ij} \leq d_2$), the multipolar (MP) approximation provides a very efficient and accurate solution of the problem (see Section 11.2.4).
The DFP and MP approximations are complementary techniques, in the sense that the former is particularly accurate for treating *close-by* pairs, whereas the latter can be safely applied only when the two WF-PAO product distributions are not overlapping, which requires the respective WFs to be distant from each other.

5.1.1 Exact two-electron integrals

rec		variable	value	meaning
•		**SCHWARZ**	**KT**	Schwarz screening for 2-el integrals is activated
•	3I	TS1 TS2 TS3	[10 10 10]	Schwarz tolerances (10^{-TS})
•		**DSCREEN**	**KT**	Pre-screening on the coefficients for 2-el integrals
•	3I	TD1 TD2 TD3	[4 4 4]	Density screening tolerances (10^{-TD})
•		**ASYMDOM**	**KT**	Only vertical excitations are considered

5.1.2 Multipolar Expansion

ERIs associated to *distant* pairs are evaluated via the so-called MP approximation: that is, the WF-PAO *product distribution* associated to either of the two electrons is described by a set of multipoles up to a maximum order ℓ, set by default to 4 and modifiable via the **MULTIPO** keyword. For further details, see Section 11.2.4.

rec	variable	value	meaning
•	**MULTIPO**	**KO**	Multipole calculation of 2-electron integrals for $d_1 \leq d \leq d_2$
• I	NMULTIP	[4]	Maximum multipole moment
•	**MPREAD**	**KT**	Read multipoles from fortran unit 97

5.1.3 Density Fitting

For the DF theory in Cryscor, see references [11, 12, 13]. The presence in the Cryscor input file of the **DFITTING** keyword, specifies that 2-electron integrals of *close-by* WF-pairs are estimated via the DF approximated technique and not by means of an exact computation.

Following **DFITTING**, a card specifying either **DIRECT** or **KSPACE** must be present. This allows to choose between the two different DF schemes present in the code. The first scheme is a fitting performed purely in the direct space, in a way formally similar to the molecular density fitting formulations. The second is a mixed scheme allowing some fitting functions to be treated with reciprocal space techniques.

In many cases the **DIRECT** scheme, with an adequate fitting basis set and default options, is sufficiently accurate and efficient.

Figure 5.1: Periodic table of elements. Red elements are those whose corresponding fitting basis set is inserted in the Cryscor library; for them, at least PG-VTZ and PG-V5Z-quality basis sets are guaranteed.

If **KSPACE** is chosen, a card must follow which specifies the shrinking factors `IS123A` and `IS123B` of the Monkhorst net used for the required Fourier Transforms (FT). The values adopted

for `IS123A` and `IS123B` have effects both on accuracy and computational times. Suggested values are: for `IS123A` the same as employed in CRYSTAL HF calculation, for `IS123B` an increased value with respect to `IS123A` (12, for instance if 8 is the HF shrinking factor). In this way, it is implicit that the same shrinking factor is used for the different directions of the reciprocal lattice. If one wants different values to be used, just set `IS123A` and/or `IS123B` to zero, and introduce 1 or 2 cards specifying the values `IS1A IS2A IS3A` and/or `IS1B IS2B IS3B`. The **MIXEXP** optional keyword defines the partition between direct and reciprocal space treatment.

General structure of the DF input block			
rec	variable	value	meaning
•	**DFITTING**	**KO**	DF for the evaluation of 2-electron integrals for *Strong* pairs
•	**DIRECT** or **KSPACE**	**KO**	Selects the DF scheme
	*If **KSPACE** then insert*		
• 2I	IS123A IS123B		Shrinking factors of **k** net in reciprocal space
•	**MIXEXP**	**KO**	Sets minimum exponent of functions in reciprocal space. All functions with lower exponent are moved in the direct space part
• F	EXP	[1.]	Exponent value
	*with either **KSPACE** or **DIRECT** insert*		
•	**PRINTBAS**	**KO**	Prints in the output the fitting basis set
•	**USRBASIS** or LABEL	**KO**	The fitting basis set is chosen; for instance `LABEL` could be `PG-VTZ`.
	*if **USRBASIS** then define the DF basis set for each type of atom*		
• 2I	AN NFF		Atomic number; number of fitting functions
	for each fitting function (F=1,NFF)		
• 3I	LAT NG T		`LAT` is the shell type, `NG` is the number of primitive functions (see the CRYSTAL Manual, Section 1.2 [3]); `T=0` if the fitting function is a GTF, `T=1` if the fitting function is a PTF
	for each primitive function (P=1,NG)		
• 2F	EXP COEF		Exponent and coefficient of primitive functions
• 2I	99 0		To end the sequence of atoms
	Optional keywords		
•	...		
•	`SEE BELOW`		
•	...		
•	**ENDDF**		To end the DF input block

Then the fitting basis set must be chosen. This can be of two types: either comprising only Gaussian Type Orbitals (GTO), or a few GTOs plus Poisson Type Orbitals (PTO). It is possible, and recommended when possible, to use one of the fitting basis sets available in the CRYSCOR

internal database (which are the molecular fitting sets used in MOLPRO [1], or the same but slightly modified in some cases [2]): basis sets of the database can be defined from input through a label composed of two parts; the first can be either `G` or `PG` depending on the character of the functions used (pure GTOs or mixed, respectively); the second part defines the quality of the set. Mixed sets include a whole set of PTOs plus one GTO per angular momentum, that is *1p 1d 1f (1g)*.

The available types of basis sets are presently:

- `G-VDZ`, `G-VTZ`, `G-VQZ`, `G-V5Z`
- `PG-VDZ`, `PG-VTZ`, `PG-VQZ`, `PG-V5Z`, `PG-AVTZ`, `PG-AV5Z`

If **DIRECT** is used, both `G` or `PG` sets can be adopted; if **KSPACE** is chosen, only `PG` sets can be adopted.
For the time being, these basis sets have been inserted into the database for light elements only (those reported in red in Figure 5.1.3) but work is still in progress.

As an alternative to predefined fitting basis, user-specified basis sets can be explicitly defined in the input by putting the `USRBASIS` flag instead of the basis set definition label, followed by the basis set in a suitable format (see the following table).

After the basis set information, further KWs for the Density Fitting can be optionally specified (see the "DF Optional Keywords" table). The default tolerances for the Schwarz selection of non negligible integrals are [8 8 16]. To change these values, the **DFSCHW** keyword must be activated. Note that when **DFITTING** is active, the tolerances defined by the KW **SCHWARZ** are ignored. The most useful command is probably the keyword **NMINCENT** , which sets the minimum size of local fit domains, which can be increased to improve the accuracy of the Density Fitting approximation, or reduced to reduce the computational weight of the calculation. These local fit-domains are defined by the program according to a shell population criterion, and symmetry adapted, so the actual size of an individual domain can be larger than defined. The default value of this parameter is different in the two fitting schemes: 12 for **DIRECT** and 6 for **KSPACE**.

Finally, the user can ask the program to print in the output file the adopted fitting basis set by means of the **PRINTBAS** keyword.

The **DFITTING** environment is closed by the **ENDDF** key. The simplest DF input blocks will then look like:

```
DFITTING
DIRECT
PG-VTZ
ENDDF
```

```
DFITTING
KSPACE
8 12
PG-VTZ
ENDDF
```

[1] www.molpro.net

[2] for instance, for some elements, when PTF sets were not available, these have been obtained by converting to PTFs the corresponding GTF set, and adding some suitable GTO functions

DF Optional keywords				
rec		variable	value	meaning
•		**DFSCHW**	**KO**	DF integrals Schwarz screening
•	3I	TS1 TS2 TS3	[8 8 16]	DF Schwarz tolerances (10^{-TS})
•		**NCONSTR**	**KT**	Multipole constraints for the local part of the fitting
•	I	NC	[0]	Number of constraints
•		**THRSHEIG**	**KT**	Sets threshold for removal of redundancies in both schemes of fitting
•	F	THR	[10^{-4}]	Threshold
*if **DIRECT** then you could insert*				
•		**METHDIR**	**KT**	Sets domains for the direct space fitting set
•	A	LABEL	[POPCENT]	Can be either `RADIUS`, `DENSITY`, `POPCENT` or `ONECENT`
•		**NMINCENT**	**KO**	Sets the minimum number of centers for either method `DENSITY` or `POPCENT`
•	I	MNCT	[12]	Number of centers
*if **KSPACE** then you could insert*				
•		**METHDIR**	**KT**	Sets domains for the direct space fitting set
•	A	LABEL	[POPCENT]	Can be either `RADIUS`, `DENSITY`, `POPCENT` or `ONECENT`
•		**METHREC**	**KT**	Sets how the fitting domains are defined for the direct space basis set
•	A	LABEL	[RADIUS]	Can be either `RADIUS`, `DENSITY`, `POPCENT` or `ONECENT`
•		**NMINCENT**	**KO**	Sets the minimum number of centers for either method `DENSITY` or `POPCENT`
•	I	MNCT	[6]	Number of centers
•		**FITRAD**	**KO**	Sets the maximum radius for fitting functions in reciprocal space (`RADIUS` method)
•	I	DISTFF	[10.]	Radius in Å

5.2 MP2 Equations

In order to better understand the meaning of the keywords commented on in this section, the reader is addressed to Section 11.2.2 of this Manual.

rec	variable	value	meaning
•	**CONTOL**	**KO**	Convergence criteria for the LMP2-SCF
• I, F	MAXC DIFF	[10 10^{-6}]	Maximum number of cycles; convergence is reached when the energy difference between two subsequent cycles is lower than `DIFF`
•	**TOLEQPRS**	**KT**	Prescreening in the evaluation of the β terms in the LMP2 equations (see Section 11.2.2)
• F	TEQ1	[10^{-8}]	First threshold for prescreening
•	**TOLEAMPS**	**KT**	Prescreening for evaluating the updates in the LMP2 equations (see Section 11.2.2)
• F	TEQ2	[10^{-8}]	Second threshold for prescreening

5.3 Restart keywords

INTREA and **AMPREA** are two *restart* KWs for the recovery of information from previous calculations. Suppose you have performed a (d_1, d_2)=(4.,8.) calculation (**ref**), which has generated data on the external units 84 and 151.
In a subsequent (4.,12.) or (6.,12.) calculation performed for the rest with the same parameters, you may want to extend the range of distant pairs (by increasing d_2), and/or to extend the range of close-by pairs (by increasing d_1) without recalculating the NPAIR integrals corresponding to the pairs classified as 'close-by' in **ref**. INTREA permits those integrals to be recovered from unit 84.
Suppose instead that you want to extract information on the MP2 density matrix corresponding to **ref**, without repeating the calculation. This can be done by inserting in the input of **ref**, both **INTREA** (in this case NPAIR is the total number of classified WF pairs) and **AMPREA** (which recovers all MP2 amplitudes from unit 151), plus the DENSMAT card (see Section 7).

rec	variable	value	meaning
•	**INTREA**	**KO**	Integral recovery from external unit 84
• I	NPAIR		Number of Flower-Flower pairs whose integrals are to be recovered
•	**AMPREA**	**KO**	Amplitudes recovery from external unit 151

Chapter 6

Dual Basis Set

In the study of extended systems, the use of diffuse basis functions is limited by the onsetting of linear dependencies in the overlap matrix. On the other hand, a correct description of the virtual manifold, often achieved with extended basis sets, plays a crucial role in electron correlation calculations. To overcome this problem a *dual basis set* option has been implemented, in which two different basis set are used; the first one, referred to as **ref**, is used for the calculation of the HF reference solution (that is to generate the WFs); the second one, including as many diffuse functions as needed, referred to as **mod**, is used in the correlation calculation (to generate PAOs and to solve the MP2 equations). The procedure is the following:

1. A CRYSTAL+PROPERTIES calculation with the **ref** basis set has to be performed in order to obtain the reference WFs (fortran units 9 and 80);

2. A CRYSTAL calculation (that reads the unit 9 of the previous calculation, renamed as unit 20) with the **mod** basis set and with the **GUESDUAL** keyword [3] frozen at the zero cycle generates the quantities of interest in the new basis set (fortran units 9 and 78);

3. A CRYSCOR calculation with the **DUALBAS** keyword has to be performed. It is mandatory to put the **DUALBAS** keyword in CRYSCOR **before** the **KNET** one.

In order to run a job with this option, the following fortran units have to be provided:
Fortran unit 80 of the **ref** calculation;
Fortran units 9 and 78 (renamed *dualbas.info*) of the **mod** calculation.

rec	variable	value	meaning
•	**DUALBAS**	**KO**	Activates the dual basis set option

Chapter 7

LMP2 correction to the HF one-electron Density Matrix

The **DENSMAT** keyword activates the MP2 correction to the HF Density Matrix calculated owing to a Lagrangian approach [14]. At present, orbital relaxation effects are not included.
The MP2-corrected valence part of the Density Matrix (DM) is saved on the external and unformatted fortran unit 63.

rec	variable	value	meaning
•	**DENSMAT**	**KO**	LMP2 Lagrangian orbital-unrelaxed Density Matrix
•	**TOLDENS**	**KO**	Tolerance for evaluating the Density Matrix via the `DENSMAT` keyword
• F	TOL	[10^{-8}]	Value of the tolerance

In order to compute any MP2-corrected property which depends on the DM (electron charge and momentum densities, Mulliken populations, structure factors, directional Compton profiles, autocorrelation function, etc.), one has to provide unit 63 to the PROPERTIES code.
By default the PROPERTIES program adopts the HF Density Matrix; but if one inserts the **PMP2** keyword followed by a number, 0 or 1, the calculated properties are referred to the whole (HF+MP2) or to the correlation only (MP2) Density Matrix, respectively. The latter option is recommended for calculating the correlation contribution to electron charge density or to Mulliken populations.
See the PROPERTIES Manual for further details on the usage of the **PMP2** keyword [3]. Let us comment, for the sake of clarity, the following PROPERTIES input:

`NEWK`	Must be inserted in the first record
`12 12 12`	New shrinking factor
`1 0`	Fermi energy must be computed
`PPAN`	(HF) Mulliken's populations are computed
`XFAC`	(HF) Structure factors are computed
`1`	(how many factors)
`2 0 0`	(which ones)
`PMP2`	Reads the MP2 correction to the Density Matrix
`0`	The HF+MP2 Density Matrix will be used
`PPAN`	(HF+MP2) Mulliken's populations are computed
`XFAC`	(HF+MP2) Structure factors are computed
`1`	(how many factors)
`2 0 0`	(which ones)
`END`	End of the input

Table 7.1: A simple example of PROPERTIES input where the **PMP2** keyword is used.

Chapter 8

Printing and Plotting

It may be useful to obtain additional printed or graphical information on some quantities of interest. Here are some suggestions for the use of the printing option activated by the **PRINPLOT** keyword.

The first argument of this keyword is the general level of printing `IPRT`. There are two main optional levels of printing: the zeroth level (0) that activates the printing of basic information (the default choice), and the first level (1) which introduces further information for experienced users. Printing levels beyond 1 are not formatted and their use is discouraged for non-developing users. Detailed information on some key-quantities of the LMP2 method can be collected if one or more of the following KWs (general name `ITEM`) is activated after the `IPRT` card (and independently of the value of IPRT): `WF FWF PAO SPAO FPAO LON SLON FLON` as illustrated below. An `ITEM=END` card must be inserted to end the printing section of the input, either immediately after the `IPRT` card, or after a group of the other `ITEM` cards.

These `ITEM` cards stand for:

- `WF` The matrix of the WF coefficients in the AO basis ($c^W[\mu\mathbf{g}; i\mathbf{0}]$);
- `FWF` The Fock matrix in the basis of the WFs ($F^W[i\mathbf{g}; j\mathbf{0}]$);
- `PAO` The matrix of the PAO coefficients in the AO basis ($c^P[\mu\mathbf{g}; a\mathbf{0}]$);
- `SPAO` The overlap matrix in the basis of the PAOs ($S^P[a\mathbf{g}; b\mathbf{0}]$);
- `FPAO` The Fock matrix in the basis of PAOs ($F^P[a\mathbf{g}; b\mathbf{0}]$);
- `LON` The matrix of the LON coefficients in the AO basis for each WF-pair ($c^L_{(ij)}[\mu\mathbf{g}; a']$);
- `SLON` The overlap matrix in the basis of LONs, for each WF-pair ($S^L_{(ij)}[a'; b']$);
- `FLON` The Fock matrix in the basis of LONs, for each WF-pair ($F^L_{(ij)}[a'; b']$).

Note that the first seven matrices (whose first index is associated to a lattice vector $\mathbf{g}$ while the second refers to the zero cell), are printed in a sequence for the different selected vectors: $\mathbf{g} = \mathbf{g_1}, \mathbf{g_2}, \ldots, \mathbf{g}_{\mathrm{NG}}$. For instance, first $\{c^W[\mu\mathbf{g} = \mathbf{g_1}; i\mathbf{0}]\}$ (for all μ's in the rows, all i's in the columns), next $\{c^W[\mu\mathbf{g} = \mathbf{g_2}; i\mathbf{0}]\}$, etc.

rec	variable	value	meaning
•	**PRINPLOT**	**KO**	Defines printing options
• I	IPRT	0	General level of printing
• A	ITEM	**KO**	`ITEM` can be one of the following keywords: **END WF FWF PAO SPAO FPAO LON SLON FLON**
	if `ITEM = FWF, SPAO, FPAO, SLON, FLON`*, then insert*		
• I	NG		Number of **g**-vectors (crystal cells) to be considered for printing
• *NG*I	NG1 NG2 $\cdots$		Labels of the selected crystal cells
	if `ITEM = WF, PAO, LON`*, then insert*		
• 2I	NG NS		Number of **g**-vectors (crystal cells) to be considered for printing and number of stars of **g**-vectors to be considered for plotting
• *NG*I	NG1 NG2 $\cdots$		Labels of the selected crystal cells
	if `ITEM = END`*, then end print section*		
	if `NS`$\neq$ *0 then insert* **MAPNET** *cards (see the* CRYSTAL *User's Manual [3])*		
• I	NPY		Number of points on the B-A segment.
	enter a keyword to choose the type of coordinates		
•	**COORDINA**	**KO**	
• 3F	XA,YA,ZA		Cartesian coordinates of point A
• 3F	XB,YB,ZB		Cartesian coordinates of point B
• 3F	XC,YC,ZC		Cartesian coordinates of point C
	or		
•	**ATOMS**	**KO**	Defines printing options
• I	IA		*Label* of the atom at point A
3I	AL,AM,AN		Indices of the cell where the atom is located
• I	IB		*Label* of the atom at point B
3I	BL,BM,BN		Indices of the cell where the atom is located
• I	IC		*Label* of the atom at point C
3I	CL,CM,CN		Indices of the cell where the atom is located
•	**END**		To end the **MAPNET** input block

Some of these quantities (WFs, PAOs and LONs) can also be sampled in a grid of points in order to be plotted in a 2D representation; in this case, data are saved on unit 25 and can be processed via the CRGRAPH graphical package provided as an external utility with the CRYSTAL program.

8.1 Timing information

If the **DTIME** keyword is inserted in the input, detailed information about the time needed to perform the different tasks of the calculation is printed in output.

rec	variable	value	meaning
•	**DTIME**	**KO**	Prints information about timings

Chapter 9

Possible sources of errors

In this section we report a brief list of possible sources of errors or mistakes that can cause wrong results:

- The quality of the MP2 correction to the HF energy and density matrix is affected by the quality of the reference HF solution. In particular, some truncations introduced at HF level and which have only small effects on the HF energy, could lead to a significant loss of correlation energy at MP2 level; this is the case, for some systems at least, of the truncation thresholds which enter the calculations of the HF two-electron integrals (governed by the **TOLINTEG** keyword); please refer to the **TOLMP2** keyword of **Crystal** in order to improve the quality of such thresholds;

- The quality of Wannier Functions is fundamental for the stability of the LMP2 procedure. In particular symmetry properties of WFs must be fulfilled with a good numerical precision. Quality of WFs can be improved *e.g.* by increasing the k-points mesh in the localization procedure. For further details refer to the PROPERTIES documentation.

- As the quality of WFs is important, so is the quality of PAOs. In particular when the dual basis set technique is used, it is important that the k-mesh for the PAO generation (**KNET** keyword) is sufficient (the energy of the singles is specially sensitive to that). If you have any doubt about the correctness of the results try increasing this parameter.

- A delicate parameter is d_1 which controls the partition of the occupied space and in particular sets the distance beyond which the multipolar expansion is used in the computation of the integrals; its value can be modified using the **PAIR** keyword. Low values of d_1 (i.e. d_1 lower than 6) may lead to inaccuracies in the computed correlation energy.

- When a binding energy $\Delta E = E(A+B) - [E(A) + E(B)]$ has to be computed, attention must be paid on the consistency of the size and shape of the excitation domains in the different calculations involved;

- The same attention as that of the previous item has to be paid when evaluating the BSSE (Basis Set Superposition Error) through the *counterpoise correction*.

- Some care should be taken when using the **MOLPAIR-ENVPAIR-MOENPAIR** set of keywords for describing physisorption phenomena: when the influence of one subsystem on the electronic cloud of the other is significant, i.e. when the molecule and substrate are close to each other, this approximation can lead to an overestimation of the interaction energy. One should then check the convergence of the energy with an increasing number of intra-pairs included, so for instance including the intra-molecule pairs and the strong pairs in the slab.

Chapter 10

What the output looks like

In this Chapter we will present and discuss the main features of the CRYSCOR output by means of some explicit examples. There are two main levels of printing among which the user can choose: the zeroth level (0) that activates the printing of basic information (which constitutes the default choice and is used in the example below) and the first level (1) which introduces further information for experienced users. The level of printing can be modified via the keyword **PRINPLOT** (see Section 8). Printing levels over 1 are not formatted and their use is discouraged for non-developing users.

10.1 The simple case of the LiH crystal

`KNET`	Construction of the HF Density Matrix (DM)
`8`	New shrinking factor
`MEMORY`	Memory required
`4000`	Value in Mbytes
`DFITTING`	Density Fitting input block
`DIRECT`	Direct-space technique
`PG-VTZ`	DF auxiliary basis set
`ENDDF`	End Density Fitting input block
`DOMDEF`	Definition of the excitation domains
`1`	There is only one flower
`1 2`	For the first flower, two stars of neighbors are included
`1 2`	The stars are the first and the second
`DENSMAT`	The MP2 correction to the HF DM is activated
`END`	End of CRYSCOR input

Table 10.1: The CRYSCOR input used for the calculation we are commenting on.

In this Section we will comment on the output of an MP2 calculation for the Lithium hydride (LiH) crystal. LiH is an ionic crystal of cubic symmetry (rocksalt structure, space group $Fm\overline{3}m$),

with only one lithium and one hydrogen per cell. The basis set adopted for Li comprises two s and one p shells while for H, three s and one p shells. This is a very simple system which allows us to comment on specific aspects of the output.

The CRYSCOR input file here adopted is the one reported in Table 10.1; it is an irreducible input apart from the **DENSMAT** keyword which activates the computation of the MP2 correction to the HF Density Matrix. The excitation domains are defined explicitly in input by means of the **DOMDEF** keyword.

10.1.1 The Header and the recovery of the information from Crystal

```
*****************************************************************************
*                                                                           *
*                                CRYSCOR09                                  *
*                                                                           *
*          An ab initio program for electron correlation in solids          *
*                                                                           *
*                           www.cryscor.unito.it                            *
*                                                                           *
*                               Main Authors                                *
*                                                                           *
*                C. PISANI(1), S. CASASSA(1), L. MASCHIO(1)                 *
*                       M. SCHUETZ(2), D. USVYAT(2)                         *
*                                                                           *
*                     (1) THEORETICAL CHEMISTRY GROUP                       *
*                   UNIVERSITA' DI TORINO - TORINO (ITALY)                  *
*            (2) INSTITUTE FOR PHYSICAL AND THEORETICAL CHEMISTRY           *
*               UNIVERSITAT REGENSBURG - REGENSBURG (GERMANY)               *
*                                                                           *
*             WHEN PUBLISHING ANY RESULT OBTAINED WITH CRYSCOR09            *
*               THE PROGRAM HAS TO BE CITED IN THE PROPER WAY:              *
*                                                                           *
*         - C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schuetz          *
*               and D. Usvyat, J. Comput. Chem., 29, 2113 (2008)            *
*                                                                           *
*               - A. Erba and M. Halo, CRYSCOR09 User's Manual              *
*           Universit degli Studi di Torino, Torino - Italy (2009)          *
*                           www.cryscor.unito.it                            *
*                                                                           *
*       THE DENSITY FITTING MODULE OF CRYSCOR09 IS BASED UPON LICENCE       *
*       ON THE CORRESPONDING MODULE OF MOLPRO (www.molpro.net)              *
*       F.R.Manby, P.J.Knowles, A.W.Lloyd, J.Chem.Phys. 115,9144(2001)      *
*                                                                           *
*****************************************************************************
```

The first portion of the CRYSCOR output is the header where some basic information about the authors of the code are reported.

```
*****************************************************************************
*                                                                           *
*     SUMMARY OF THE COMPUTATIONAL SETUP OF THE HF REFERENCE CALCULATION     *
*                                                                           *
*****************************************************************************

*****************************************************************************
LiH
CRYSTAL - PROPERTIES - TYPE OF CALCULATION :  RESTRICTED CLOSED SHELL
*****************************************************************************
HARTREE-FOCK HAMILTONIAN

DIRECT LATTICE VECTOR COMPONENTS (BOHR)
        0.00000    3.66229    3.66229
        3.66229    0.00000    3.66229
        3.66229    3.66229    0.00000

LATTICE PARAMETERS  (BOHR AND DEGREES) - PRIMITIVE CELL
      A          B          C         ALPHA      BETA     GAMMA        VOLUME
   5.17926    5.17926    5.17926     60.0000   60.0000   60.0000      98.23990
*****************************************************************************
N. OF ATOMS PER CELL        2  COULOMB OVERLAP TOL         (T1) 10** -6
NUMBER OF SHELLS            7  COULOMB PENETRATION TOL     (T2) 10** -6
NUMBER OF AO               11  EXCHANGE OVERLAP TOL        (T3) 10** -6
N. OF ELECTRONS PER CELL    4  EXCHANGE PSEUDO OVP (F(G))  (T4) 10** -6
CORE ELECTRONS PER CELL     2  EXCHANGE PSEUDO OVP (P(G))  (T5) 10**-12
N. OF SYMMETRY OPERATORS   48  POLE ORDER IN MONO ZONE                4
*****************************************************************************
  ATOM N.AT.  SHELL   X(AU)     Y(AU)     Z(AU)     EXAD         N.ELECT.
*****************************************************************************
   1    1 H     4     0.000     0.000     0.000   3.000E-01       1.983
   2    3 LI    3     3.662     3.662     3.662   6.000E-01       2.017
*****************************************************************************
DE(K)                    0.000E+00  ENERGY LEVEL SHIFTING        0.00000
TOTAL ENERGY -8.0592434161786E+00  CONV. ON ENERGY            3.742E-08
KIN. ENERGY   8.1334688248138E+00  VIR. COEFF.           1.00458388E+00
N. OF SCF CYCLES                8  FERMI ENERGY              -0.229E+00
WEIGHT OF F(I) IN F(I+1)       30
SHRINK. FACT.(MONKH.)     8  8  8  SHRINKING FACTOR(GILAT NET)        8
NUMBER OF K POINTS IN THE IBZ  29  CELL VOLUME (A.U.)          98.23990
*****************************************************************************
```

The second part reports a summary of the main computational parameters used and of the main structural features of the reference HF calculation, performed with CRYSTAL.

10.1.2 The recovery of the information from Properties

After the recovery of the information coming from the HF calculation (read from the external fortran unit 9), the CRYSCOR program reads from the external unit 80 the information concerning the localization and symmetrization of the crystalline orbitals into Wannier functions (WF). The symmetry properties of such functions is reported in the output:

```
*****************************************************************************
*                                                                           *
*      WANNIER FUNCTIONS (WF) INFORMATION FROM THE PROPERTIES CALCULATION    *
*                                                                           *
*****************************************************************************


-----------------------------------------------------------------------------
BUNCH  FLOWER  FLOWER  PETAL |    TYPE         ATOM1  (G1)  -   ATOM2  (G2)
         ABS     REL         |
-----------------------------------------------------------------------------
   1       1       1      1  |  NON BONDING    H  1   ( 1)  -
-----------------------------------------------------------------------------
```

The case of LiH is very simple in the present respect, since it has only two valence electrons per cell, hence only one WF (of atomic type) centered in the hydrogen atom, essentially of s character.

10.1.3 The computational setup of the MP2 calculation

In this section of the output, the number of AOs and WFs per unit cell is first reported, followed by a list of the most important CRYSCOR parameters adopted in the calculation; some of them may be explicitly defined in input while others assume their default values.

The maximum number of cycles for the solving procedure of the MP2 equations and the corresponding converge threshold are reported (they can be modified via the **CONTOL** keyword, see Section 5.2). The meaning of the pair distances (that can be modified via the dedicated **PAIR** keyword) will be discussed in depth below. The truncation thresholds for the expansion of WFs and PAOs in the AOs basis, are set to 10^{-4} and can be modified via the **TOBJ** keyword. The adjoined GTF exponent used in the recalculation of the one-electron Density Matrix for generating the PAOs is set to 0.04, which corresponds to high accuracy. The Density Fitting technique is used in the calculation of some two-electron integrals; excitation domains are explicitly defined in input and the global level of printing is 0 (default). Other parameters are more technical (see the previous Chapter for more details).

```
*****************************************************************************
*                                                                           *
*                  ---> THE MP2 CALCULATION STARTS HERE <---                *
*                            (COMPUTATIONAL SETUP)                          *
*                                                                           *
*****************************************************************************


  NUMBER OF AOs:    11
  NUMBER OF WFs:     1

  INPUT PARAMETERS:

  MAXIMUM NUMBER OF CYCLES:                        30
  CONVERGENCE ON ENERGY:                    1.0000E-06
  PAIR DISTANCE: WEAK                       8.0000E+00
                 DISTANT                    1.2000E+01
  PAOs-COEFFICIENTS TOLERANCE:              1.0000E-04
  WFs-COEFFICIENTS  TOLERANCE:              1.0000E-04
  ADJOINT GTF EXPONENT FOR P TRUNCATION     4.0000E-02
  SCHWARZ TOL STRONG WEAK DISTANT: 10-       8   8  16
  DSCREEN TOL STRONG WEAK DISTANT: 10-       4   4   4
  MONKHORST SHRINKING FACTORS:               0   0   0
  LEVEL OF PRINTING:                                 0
  DENSITY FITTING

  MULTIPOLAR CORRECTION STARTS AT (A):      8.0000E+00
  EXCLUDE PAOs WITH NORM LESS THAN:10-               6
  DOMAINS DEFINED BY INPUT
```

10.1.4 The operating region

Some quantities which define the operating region of the calculation are then printed. The number of unit cells defining the active region of the crystal is reported along with the number of **k**-points used in the reciprocal space in order to generate the PAOs.

```
*****************************************************************************
*                                                                           *
*                              OPERATING REGION                             *
*                                                                           *
*****************************************************************************
```

```
--->  19237 CRYSTAL CELLS (   204 STARS OF G-VECTORS )
--->    MAX DENSITY FITTING RANGE :          600
--->   PAOs ARE GENERATED IN THE RECIPROCAL SPACE
--->    512 K-POINTS IN RECIPROCAL SPACE
--->    THE LARGEST WF-PAO OVERLAP IS:    0.38581383E-09
```

10.1.5 The excitations domains

Information on the size and shape of WF excitation domains is next reported. Since all petals in the same flower have the same domain (because each domain is symmetrized according to the invariant subgroup of symmetry of the corresponding flower) this information only concerns the different flowers.

```
*****************************************************************************
*                                                                           *
*                          FLOWERS EXCITATION DOMAINS                       *
*                                                                           *
*****************************************************************************


DEFINED AS THE SET OF PAOs BELONGING TO THE FOLLOWING SET OF ATOMS:

FLOWER N.  1

ATOM  1 H    -   N. OF G:   1  --->    1
ATOM  2 LI   -   N. OF G:   6  --->   12   8   2  40  30  26
```

Here, the **DOMDEF** keyword has been used. From the input shown in Table 10.1, we see that just one domain is defined: indeed, there is just one flower in the system to which two stars of neighbors are assigned; they are specified in the following input card: star 1, which contains one H atom and star 2, which contains six Li atoms, for a total of 7 atoms. The list of the atoms of either type in the domain, and the crystal cell to which they belong are reported in the output.

10.1.6 The kernel domains

A *minimal* domain $[i0]^0$ can be associated to each reference WF $\phi_{i0}(\mathbf{r})$ and then transferred to the whole set $\{\phi_i(\mathbf{r})\}$ of its periodic images in the crystal. This minimal domain is defined as the set of any atom of the system contributing to the total Mulliken electronic population of the WF $\phi_{i0}(\mathbf{r})$ for more than a given threshold Q^0 (see Section 11.1.4). If Q^0 is reasonably chosen then the *minimal* domain represents a minimal chemical description of the corresponding WF. If the selected WF is a *bond* one then the corresponding *minimal* domain will be the union of the two atoms belonging to the bond; otherwise if the selected WF is an *atomic* one, then the corresponding *minimal* domain will contain only one atom (the one in which the WF is centred).

Minimal domains define the metric of WF-pairs.

```
*****************************************************************************
*                                                                           *
*                            FLOWERS KERNEL DOMAINS                         *
*                                                                           *
*****************************************************************************


DEFINED AS THE SET OF PAOs BELONGING TO THE FOLLOWING SET OF ATOMS:

FLOWER N.  1

ATOM  1 H    -   N. OF G:   1  --->    1
```

In the case of LiH, since the only flower is of atomic type, its minimal domain is just made up of one H atom.

10.1.7 The pairs partition (distance)

As previously said, the **PAIR** keyword serves to classify WF-pairs according to the distance between the corresponding minimal domains. Two distances d_1 and d_2 are defined (the default values being 8 and 12 Å, respectively).

This classification concerns in fact pairs of flowers, since the domains of all WFs in the same flower coincide. Furthermore, only the symmetry irreducible pairs are classified since symmetry is fully exploited.

```
*****************************************************************************
*                                                                           *
*                       PAIRS CLASSIFICATION (DISTANCE)                     *
*                                                                           *
*****************************************************************************


NUMBER OF STRONG  PAIRS:        1
NUMBER OF WEAK    PAIRS:        8
NUMBER OF DISTANT PAIRS:       15

NUMBER OF IRREDUCIBLE PAIRS CLASSIFIED:       24
```

There is just one *Strong* pair ($d = 0$) while there are 8 *Weak* pairs ($0 < d \leq d_1$) and 15 *Distant* pairs ($d_1 < d \leq d_2$), for a total of 24 irreducible pairs.

Pairs separated by more than d_2 are considered *Very distant*, and are neglected in the local approach.

It is important to state that the classification of the pairs according to the distance is strictly related to the way the respective integrals are treated. Concerning this point, it can be observed

that *Strong* and *Weak* pairs are always treated in the same approximation, so that this distinction is purely formal for the moment being. In the present case, they are calculated with the Density Fitting technique.

On the contrary, the integrals of *Distant* pairs are evaluated via a multipolar approximation.

10.1.8 Two-electrons Integrals

The two-electron integrals which correspond to *Strong* and *Weak* WF-pairs are treated via the Density Fitting approximation. Information on the time required for such part of the calculation is reported:

```
*****************************************************************************
*                                                                           *
*         INTEGRALS CALCULATION VIA THE DENSITY FITTING APPROXIMATION       *
*                                                                           *
*****************************************************************************


TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT DFITINIT    TELAPSE       21.77 TCPU       20.14
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT MP2INTEG    TELAPSE      329.99 TCPU      325.58
```

The integrals corresponding to *Distant* WF-pairs are then evaluated via the so-called Multipolar approximation: multipoles (up to hexadecapoles by default) are first constructed and then integrals are computed.

```
*****************************************************************************
*                                                                           *
*                              MULTIPOLAR INTEGRALS                         *
*                                                                           *
*****************************************************************************


TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT MPOLESCALC  TELAPSE       19.58 TCPU       19.57
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT MPOLESINTE  TELAPSE       19.69 TCPU       19.69
```

10.1.9 LMP2 equations

The LMP2 equations are then solved through a self consistent procedure. At each cycle, the total correlation energy (ECORR) of the system is printed; this energy does not include the contribution of the residues (DELTA) that must vanish at convergence.

```
*****************************************************************************
*                                                                           *
*                      ITERATIVE SOLVER OF THE LMP2 EQUATIONS               *
*                                                                           *
*****************************************************************************


CYCLE  1     ECORR  -0.02810152     DELTA -0.00007950
CYCLE  2     ECORR  -0.02815544     DELTA -0.00005392
CYCLE  3     ECORR  -0.02815541     DELTA  0.00000089
```

The total absolute value of the residues is also given at each cycle. When the residues become smaller than the convergence threshold set by the **CONTOL** keyword (10^{-6} in this case) the procedure stops and convergence is reached:

```
*****************************************************************************
*                                                                           *
*             MP2 CONVERGENCE REACHED --->  E(MP2):  -0.02815541            *
*                                                                           *
*****************************************************************************
```

This is the final MP2 correlation energy of the crystalline system per cell.

10.1.10 The Grimme correction

Grimme has recently proposed [Grimme, *J. Chem. Phys.*, **118**, 9095 (2003)] a semi-empirical scaling of the two spin components of singlet and triplet. This strategy has been implemented in the CRYSCOR code. The total correlation energy obtained within such a scheme, along with its two components, is then reported in the output.

```
*****************************************************************************
*                                                                           *
*                                        E(GRIMME):  -0.03101104            *
*                                  GRIMME COMPONENTS:                       *
*                                       ---> TRIPLET:  -0.00320244          *
*                                       ---> SINGLET:  -0.02495297          *
*                                                                           *
*****************************************************************************
```

10.1.11 Singly excited configurations

According to the Brillouin theorem, singly excited configurations should not contribute in any extent to the total correlation energy of a system. This condition should be verified when calculations are performed using the same basis set for the HF reference calculation and for the LMP2 correlation calculation. However, when dealing with the Dual Basis Set option (see the **DUALBAS** keyword in Section 6) this condition is no more expected to be verified. An algorithm which permits to evaluate the contribution of singly excited configurations to the correlation energy has been implemented.

```
*****************************************************************************
*                                                                           *
*                                      E(SINGLES):   0.00000000             *
*                                 E(MP2+SINGLES):  -0.02815541              *
*                              E(GRIMME+SINGLES):  -0.03101104              *
*                                                                           *
*****************************************************************************
```

10.1.12 Final Summary

A final summary recovers the HF starting energy and adds to it the MP2 correlation energy:

```
*****************************************************************************
*                                                                           *
*                                        HF ENERGY:  -0.8059243416E+01      *
*                                   HF+MP2 ENERGY:  -0.8087398824E+01      *
*                                HF+GRIMME ENERGY:  -0.8090254458E+01      *
*                                                                           *
*****************************************************************************
```

10.1.13 The Lennard-Jones extrapolation to infinity

Taking advantage from the well-known London law r^{-6} governing the behavior of the pair-energies with respect to pair-distancies, a scheme as been implemented which permits to recover the contribution of *Very Distant* pairs, otherwise completely neglected, to the correlation energy by means of a Lennard-Jones extrapolation.

This strategy is activated by default for any 2D and 3D system. First, the parameters used during this procedure are written; the user can modify them by virtue of the **LENJONES** keyword.

```
*****************************************************************************
*                                                                           *
*          INPUT PARAMETERS FOR LENNARD-JONES CALCULATION (ANGSTROM)        *
*                                                                           *
*****************************************************************************
```

```
MIN FL-FL DIST. FOR COEFFICIENT CALCULATION (D-LJ1):     6.00
MAX DISTANT-PAIR DISTANCE                   (D-LJ2):    12.00
MAX RADIUS                                  (D-LJ3):    63.04
ENERGY TRUNCATION TOLERANCE                (TOL-LG):        0
```

The extrapolated contribution to the correlation energy is then printed along with a summary of the HF and MP2 energies:

```
*****************************************************************************
*                                                                           *
*              LENNARD-JONES EXTRAPOLATED ENERGY CONTRIBUTIONS:             *
*                                                                           *
*****************************************************************************

 LJ(MP2)                  -0.0000013065
 LJ(GRIMME)               -0.0000010032

 HF ENERGY                -8.0592434162E+00
 HF+MP2+LJ ENERGY         -8.0874001308E+00
 HF+GRIMME+LJ ENERGY      -8.0902554612E+00
```

10.1.14 MP2 correction to the HF Density Matrix

Since we put the **DENSMAT** keyword in the input, the calculation of the MP2 correction to the HF Density Matrix is activated.

```
*****************************************************************************
*                                                                           *
*                    MP2 CORRECTION TO THE HF DENSITY MATRIX                *
*                                                                           *
*****************************************************************************

 LMP2 DENSITY MATRIX CORRECTION FOR THE OCCUPIED ORBITALS
 MAXIMAL NONSYMMETRICITY      0.1901E-07
 THE TRACE IN AOs            -0.2652E-01

 LMP2 DENSITY MATRIX CORRECTION FOR THE VIRTUAL ORBITALS
 MAXIMAL NONSYMMETRICITY      0.3676E-07
 THE TRACE IN AOs             0.2652E-01

 TOTAL LMP2 DENSITY MATRIX CORRECTION
 MAXIMAL NONSYMMETRICITY      0.4078E-07
 THE TRACE IN AOs             0.6697E-07
```

The MP2 correction to the HF Density Matrix is then written on the external unformatted fortran unit 63 which can be passed to the PROPERTIES program using the **PMP2** keyword in order to compute any DM-related quantity.

```
*****************************************************************************
 MP2 DENSITY MATRIX HAS BEEN WRITTEN ON UNIT 63
*****************************************************************************
```

10.2 A more complex case: the hexagonal Argon crystal

In this Section we comment on the output of an MP2 calculation for the hexagonal structure of Argon crystal. This is a more interesting case with respect to LiH both because there are more WFs per atom and because two Ar atoms per cell are present; a richer analysis of bunches, flowers and petals can thus be carried out. Note that only relevant points of the output are commented on, in a perspective of highlighting the differences with respect to the previously fully analysed example of LiH.

10.2.1 The recovery of the information from Properties

The table collecting information on the localization and symmetrization of WFs and their classification looks as follows:

```
*****************************************************************************
*                                                                           *
*      WANNIER FUNCTIONS (WF) INFORMATION FROM THE PROPERTIES CALCULATION    *
*                                                                           *
*****************************************************************************


-----------------------------------------------------------------------------
BUNCH  FLOWER  FLOWER  PETAL |    TYPE         ATOM1  (G1)  -   ATOM2  (G2)
         ABS     REL         |
-----------------------------------------------------------------------------
   1       1       1      1  |  NON BONDING    AR 1   ( 1)  -
   1       2       2      2  |  NON BONDING    AR 2   ( 1)  -
-----------------------------------------------------------------------------
   2       3       1      3  |  NON BONDING    AR 1   ( 1)  -
   2       3       1      4  |  NON BONDING    AR 1   ( 1)  -
   2       4       2      6  |  NON BONDING    AR 2   ( 1)  -
   2       4       2      7  |  NON BONDING    AR 2   ( 1)  -
-----------------------------------------------------------------------------
   3       5       1      5  |  NON BONDING    AR 1   ( 1)  -
   3       6       2      8  |  NON BONDING    AR 2   ( 1)  -
-----------------------------------------------------------------------------
```

DUALBAS	Dual Basis Set option activated
KNET	Construction of the HF Density Matrix (DM)
8	New shrinking factor
MEMORY	Memory required
2000	Value in Mbytes
DFITTING	Density Fitting input block
DIRECT	Direct-space technique
PG-VTZ	DF auxiliary basis set
ENDDF	End Density Fitting input block
DOMDEF	Definition of the excitation domains
6	There are six flowers
1 3	For the first flower, three stars of neighbors are included
1 2 3	The stars are the first the second and the third
2 3	
1 2 3	
3 3	
1 2 3	
4 3	
1 2 3	
5 3	
1 2 3	
6 3	
1 2 3	
PAIR	Only WF pairs up to 4 Å are calculated via the DF approximation
4. 4.	
END	End of CRYSCOR input

Table 10.2: The CRYSCOR input used for the calculation we are commenting on.

Please also refer to Chapter 2 to fully comprehend this commented output part.
In each cell there are two symmetry-related Ar atoms which are the site of the SAWFs. Consider Ar 1 as the reference site. Its site-symmetry is C_{3v} (as that of N in ammonia). Its 8 valence electrons will be associated to 4 WFs, which, intuitively, will have a symmetry of type s, p_x, p_y and p_z.

The total of 8 WF/petals (4 for each Ar) is classified in 3 bunches and 6 flowers: the first bunch contains two flowers with one petal each, the second two flowers with two petals each, the third one two flowers with one petal each. What does this mean?

It means that the two s-type WF of each Ar are grouped into one bunch (with one flower, one petal), whereas the other 6 petals/WF (3 per each Ar) group is divided into two bunches: one containing, let's say, the p_z-like orbitals (2, one per each Ar), and the other one containing the rest, i.e. the p_x, p_y-like orbitals (4, two per Ar). For this bunch, corresponding to Bunch n. 2, the two flowers contain each two petals, i.e. the degeneracy p_x-p_y between WF is recognized. Evidently,

in the hexagonal Ar crystal, the p_x-p_y-p_z degeneracy does not hold any more and only the p_x-p_y is left.

Is there a way to verify these conjectures? Let us check the petal (WF) composition one by one. We need to refer to the *properties* output for this. So for instance either petals n. 1 and 2 in bunch 1 or petals n. 5 and 8 in bunch 3 have to have a prevalently s-orbitals composition.

```
~~~
```

```
 A.O. POPULATIONS OF ATOM    1(  0  0  0):
   0.0009 (-> S)
  -0.0162 (-> S)
   0.0000 (-> Px)  0.0000 (-> Py)  0.0000 (-> Pz)
   0.1862 (-> S)
   0.0000 (-> Px)  0.0000 (-> Py)  0.0000 (-> Pz)
   0.5474 (-> S)
   0.0000 (-> Px)  0.0000 (-> Py)  0.0000 (-> Pz)
   0.2809 (-> S)
   0.0000 (-> Px)  0.0000 (-> Py)  0.0000 (-> Pz)
```

Therefore petal n. 1 is effectively of s-type symmetry; so is consequently and effectively petal n.2 of the 1st bunch (orbital s-like for the second Ar atom in the cell).

```
 WANNIER NUMBER    2 ATOMIC EXTENT  1.0016
[...]
 A.O. POPULATIONS OF ATOM    2(  0  0  0):
   0.0009  -0.0162   0.0000   0.0000   0.0000   0.1862   0.0000   0.0000
   0.0000   0.5474   0.0000   0.0000   0.0000   0.2809   0.0000   0.0000
   0.0000
```

Then petals 5 and 8 are p_z-like (one per each Ar, indicated as ATOM 1 and ATOM 2) and are grouped in bunch 3.

```
 WANNIER NUMBER    5 ATOMIC EXTENT  1.0073
[...]
 A.O. POPULATIONS OF ATOM    1(  0  0  0):
   0.0000   0.0000   0.0000   0.0000  -0.0028   0.0000   0.0000   0.0000
   0.3269   0.0000   0.0000   0.0000   0.3578   0.0000   0.0000   0.0000
   0.3145
~~~
```

```
 WANNIER NUMBER    3 ATOMIC EXTENT  1.0074
[...]
 A.O. POPULATIONS OF ATOM    1(  0  0  0):
   0.0000   0.0000  -0.0028   0.0000   0.0000   0.0000   0.3268   0.0000
   0.0000   0.0000   0.3580   0.0000   0.0000   0.0000   0.3143   0.0000
   0.0000
~~~
```

```
*****************************************************************************
*                                                                           *
*                          FLOWERS EXCITATION DOMAINS                       *
*                                                                           *
*****************************************************************************


DEFINED AS THE SET OF PAOs BELONGING TO THE FOLLOWING SET OF ATOMS:

FLOWER N.  1

ATOM  1 AR   -   N. OF G:   7  --->    1   5   6   3   2   7   4
ATOM  2 AR   -   N. OF G:   6  --->   43  21  26  15   5   6

FLOWER N.  2

ATOM  1 AR   -   N. OF G:   6  --->   42  20  27  14   4   7
ATOM  2 AR   -   N. OF G:   7  --->    1   4   7   2   3   6   5

FLOWER N.  3

ATOM  1 AR   -   N. OF G:   7  --->    1   2   4   6   7   5   3
ATOM  2 AR   -   N. OF G:   6  --->   26   6   5  15  43  21

FLOWER N.  4

ATOM  1 AR   -   N. OF G:   6  --->   27   7  42  20   4  14
ATOM  2 AR   -   N. OF G:   7  --->    1   3   5   7   6   4   2

FLOWER N.  5

ATOM  1 AR   -   N. OF G:   7  --->    1   5   6   3   2   7   4
ATOM  2 AR   -   N. OF G:   6  --->   43  21  26  15   5   6

FLOWER N.  6

ATOM  1 AR   -   N. OF G:   6  --->   42  20  27  14   4   7
ATOM  2 AR   -   N. OF G:   7  --->    1   4   7   2   3   6   5
```

An alternative way would have been the use of the **DOMDEF2** keyword, in which case the input would be:

```
DOMDEF2
6
13 13 13 13 13 13
```

with an output result analogous to what just discussed. Note also that, since additional atoms are automatically added in order to fill the outermost star of atoms, the same result is obtained for other values of **DOMDEF2** comprised between 8 and 13 atoms.

10.2.3 The Dual Basis Set option

Since the Dual Basis Set is activated here, we expect the contribution of singly excited configurations to be *non* zero. This quantity should be taken into account when providing the final MP2 energy, and is automatically done in the output, as shown below:

```
                ENERGY DUE TO SINGLY EXCITED CONFIGURATIONS
               (IT SHOULD BE ZERO FOR THE BRILLOUIN THEOREM)

 ****************************************************************************
 *                                                                          *
 *                                     E(SINGLES):  -0.00026316             *
 *                                E(MP2+SINGLES):  -0.30121289              *
 *                             E(GRIMME+SINGLES):  -0.29010508              *
 *                                                                          *
 ****************************************************************************

                              FINAL SUMMARY

 ****************************************************************************
 *                                                                          *
 *                                      HF ENERGY:   0.0000000000E+00       *
 *                                 SINGLES ENERGY:  -0.2631583767E-03       *
 *                           MP2 CORRELATION ENERGY:  -0.3009497268E+00     *
 *                            HF+SINGLES+MP2 ENERGY:  -0.3012128852E+00     *
 *                     HF+SINGLES+SCS(GRIMME) ENERGY:  -0.2901050812E+00     *
 *                                                                          *
 ****************************************************************************
```

Chapter 11

Theory in pills

The difference between the "exact" ground-state energy E_0 and the Hartree-Fock (HF) energy E^{HF} of a system is a negative quantity called *correlation energy*:

$$E^{corr} = E_0 - E^{HF} \tag{11.1}$$

The ground-state wavefunction $|\Phi_0\rangle$ of an N-electron system can always be written as follows in terms of the N-detors constructed from the HF spin-orbitals (so):

$$|\Phi_0\rangle = |\Psi_0\rangle + \sum_h^a A_h^a |\Psi_h^a\rangle + \sum_{hk}^{ab}{}^{\gg} A_{hk}^{ab} |\Psi_{hk}^{ab}\rangle + \sum_{hkl}^{abc}{}^{\gg} A_{hkl}^{abc} |\Psi_{hkl}^{abc}\rangle + \cdots \tag{11.2}$$

Here $|\Psi_0\rangle$ is the HF detor, containining the N occupied so's $(h, k, l, \dots)$, $|\Psi_{hkl\dots}^{abc\cdots}\rangle$ is the detor obtained by substituting in the HF detor the so's $h, k, l, \dots$ with the virtual ones $a, b, c, \dots$; the symbol $\gg$ means that the indices must be $h > k > \cdots$ and $a > b > \cdots$. By left-multiplying equation (11.2) by $\langle\Psi_0|\hat{H}$, the following exact expression for the correlation energy is easily obtained:

$$E^{corr} = \langle\Psi_0|\hat{H}|\Phi_0\rangle - \langle\Psi_0|\hat{H}|\Psi_0\rangle = E_0\langle\Psi_0|\Phi_0\rangle - E^{HF} = \sum_{hk}^{ab}{}^{\gg} A_{hk}^{ab} \langle\Psi_0|\hat{H}|\Psi_{hk}^{ab}\rangle \tag{11.3}$$

In the last passage the Brillouin theorem has been used, along with the fact that the fundamental configuration does not *interact* with the triply (or more) excited ones. It follows that the correlation energy is perfectly determined when the coefficients of the doubly excited configurations are known.

11.1 The local approach

In this subsection we present the main features of the local correlation methods. As it is well known from the molecular experience, the computational cost of conventional electron correlation methods like MP2 or CCSD(T) increases dramatically with the size of the system. Being $\mathcal{N}$ the size of the system, the scaling of the computational effort (CPU calculation time, required memory and disk) for the MP2 method is $\mathcal{O}(\mathcal{N}^5)$. This scaling becomes $\mathcal{O}(\mathcal{N}^7)$ for MP4 and $\mathcal{O}(\mathcal{N}^8)$ for

MP5. It is clear that such a *scaling wall* avoids the usage of these canonical correlation-methods for the study of *big* systems. The steep scaling mainly originates from the delocalized character of the canonical Molecular Orbitals (MO) basis. The increase of the CPU time with the molecular size is $\mathcal{O}(\mathcal{N}^7)$ for the best method of choice, which is usually CCSD(T).

From a physical point of view, however, there should be no need to correlate all electrons in an extended molecular system (or in a crystal): dynamic electron correlation in non-metallic systems is a short range effect with an asymptotic distance dependence $\propto r^{-6}$ (dispersion energy), and thus the high-order dependence of the computational cost with the number of electrons of the system is just an artifact of the canonical orthogonal basis, in which the diverse correlation methods have traditionally been formulated. One natural way to circumvent this problem is to use *local* orbitals to span the occupied and virtual spaces. Such local correlation methods have been proposed by several authors.

Particularly successful has been the local correlation method originally proposed by Pulay [7], which was first implemented by Saebø and Pulay for Møller-Plesset perturbation theory up to fourth order (LMP2 - LMP4(SDQ) without triple excitations) [8]. While in the early works of Saebø and Pulay [8] [15] it could already be shown that only 1-2% of the correlation energy (relative to conventional calculations with the same basis set) is lost by the local approximation, it was not yet possible at the same time to demonstrate that the scaling of the computational cost can actually be reduced, and that larger systems than with conventional methods can be treated.

Significant progress in this direction was only made during the last few years when the local correlation methods were combined with newly developed integral-direct techniques [16]. Within such a framework, it has been possible to develop $\mathcal{O}(\mathcal{N})$ algorithms (asymptotic linear scaling of all computational resources, i.e. CPU time, memory and disk space) for Local-MP2 [17] and Local-CCSD.

In the local correlation methods as applied to molecules, the occupied space is usually spanned by localized molecular orbitals (LMOs), which are obtained from the occupied canonical orbitals of a previous SCF calculation by virtue of a unitary localization procedure [18] [19], which maintains the orthogonality of the occupied SCF orbitals. The idea of Pulay was to abandon the orthogonality of the virtual orbitals, and to use a basis of functions which resemble the atomic orbitals (AOs) as much as possible. Obviously, the AOs are optimally localized, but since they are not orthogonal to the occupied manifold one cannot use them straightforwardly. The strong orthogonality between the occupied and virtual spaces must be retained, since otherwise excitations would violate the Pauli exclusion principle and the theory would become very complicated. Localized virtual orbitals (PAOs) can be obtained by projecting out the occupied space from the AOs.

11.1.1 Local description of the occupied manifold: Wannier Functions

Wannier functions (WFs) are used in the CRYSCOR program to span the occupied HF manifold. They are obtained from the set of occupied crystalline orbitals (COs) via a unitary transformation and they are provided by the CRYSTAL code. WFs are quadratically integrable in $\mathcal{R}^3$ and, therefore, they are always mostly localized within a finite spatial region [5]. Localized orbitals are extensively used in molecular quantum chemical applications. They permit an easy and intuitive description of the electronic structure of the system in terms of chemical well known concepts such as lone

pairs, shared electrons, covalent or chemical bonds. Furthermore localized orbitals are a promising starting point for the treatment of electron correlation in extended systems within the so-called local methods. In the CRYSTAL program a variational basis set of contracted Gaussian-type Functions (GTFs) is used. Each GTF is the product of a real solid spherical harmonic times a radial Gaussian function. Basis functions (indicated as $\chi_\mu(\mathbf{r}-\mathbf{s}_\mu-\mathbf{T}_\mathcal{M})$) are linear combinations of primitive GTFs centered in $\mathbf{s}_\mu+\mathbf{T}_\mathcal{M}$, where $\mu=1,2,\ldots,M$ labels the basis functions assigned to the reference primitive cell; $\mathbf{s}_\mu$ is the position of a given atom within the reference cell and $\mathbf{T}_\mathcal{M}$ is a lattice vector that translates this atom to its corresponding periodic image in cell $\mathcal{M}$. By convention $\mathbf{T}_0=\mathbf{0}$ and $\mathcal{M}\oplus\mathcal{L}$ stands for $\mathbf{T}_{\mathcal{M}\oplus\mathcal{L}}=\mathbf{T}_\mathcal{M}+\mathbf{T}_\mathcal{L}$ (analogously the $\mathcal{M}\ominus\mathcal{L}$ symbol is used to label the difference of two lattice vectors). These contracted GTFs are called Atomic Orbitals (AOs).

A Wannier function in the reference cell $\phi_{i0}(\mathbf{r})$ can be expressed in terms of the AO basis set as follows:

$$\phi_{i0}(\mathbf{r})=\sum_{\mu=1}^{M}\sum_{\mathcal{M}=0}^{L-1}L_{\mu,i}^{\mathcal{M}}\chi_\mu(\mathbf{r}-\mathbf{s}_\mu-\mathbf{T}_\mathcal{M}) \tag{11.4}$$

where the first and second sums in the RHS run over the M AOs in the reference cell and the L cells of the system, respectively. It is worth noticing that L is actually ∞, but infinite sums can in general be restricted to a finite number of terms. WFs $\phi_{i0}(\mathbf{r})$ are conventionally assigned to the reference cell and hence in the following they will be called *reference* WFs. In CRYSCOR the WFs are truncated (see keyword **TOBJ** at page 23). Exploiting translational invariance one can always write:

$$\begin{aligned}\phi_{i\mathcal{I}}(\mathbf{r}) &= \phi_{i0}(\mathbf{r}-\mathbf{T}_\mathcal{I})\\ &= \sum_{\mu=1}^{M}\sum_{\mathcal{M}=0}^{L-1}L_{\mu,i}^{\mathcal{M}\ominus\mathcal{I}}\chi_\mu(\mathbf{r}-\mathbf{s}_\mu-\mathbf{T}_\mathcal{M})\end{aligned} \tag{11.5}$$

The set $\{\phi_i(\mathbf{r})\}$ of all the periodic images of the *reference* WF $\phi_{i0}(\mathbf{r})$ does additionally fulfill the following orthonormality condition:

$$\int\phi_i(\mathbf{r}-\mathbf{T}_\mathcal{I})^*\phi_i(\mathbf{r}-\mathbf{T}_{\mathcal{I}'})d\mathbf{r}=\langle\phi_{i\mathcal{I}}\mid\phi_{i\mathcal{I}'}\rangle=\delta_{\mathcal{I}\mathcal{I}'} \tag{11.6}$$

that can be generalized to any pair of Wannier functions of the system:

$$\int\phi_i(\mathbf{r}-\mathbf{T}_\mathcal{I})^*\phi_i(\mathbf{r}-\mathbf{T}_{\mathcal{I}'})d\mathbf{r}=\langle\phi_{i\mathcal{I}}\mid\phi_{i'\mathcal{I}'}\rangle=\delta_{\mathcal{I}\mathcal{I}'}\delta_{ii'} \tag{11.7}$$

where integration is performed over the whole $\mathcal{R}^3$ space. Owing to the localized nature of the AO basis set, a Mulliken analysis [20] [21] can be performed to obtain the atomic populations of the i-th WF:

$$Q_{A\mathcal{A},i0}=\sum_{\mu\in A\mathcal{A}}\sum_{\nu=1}^{M}\sum_{\mathcal{N}=0}^{L-1}L_{\mu,i}^{\mathcal{A}}\left(L_{\nu,i}^{\mathcal{N}}\right)^*S_{\mu\mathcal{A}\nu\mathcal{N}} \tag{11.8}$$

The above expression refers to the Mulliken population contribution of atom A (in cell $\mathcal{A}$) to the i-th *reference* WF. The first sum in the RHS runs over basis functions belonging to (i.e. centered on) atom $A\mathcal{A}$; $S_{\mu\mathcal{A}\nu\mathcal{N}}$ is the overlap matrix which is given by

$$S_{\mu\mathcal{A}\nu\mathcal{N}} = \int d\mathbf{r}\chi_\mu(\mathbf{r}-\mathbf{s}_\mu-\mathbf{T}_\mathcal{A})\chi_\nu(\mathbf{r}-\mathbf{s}_\nu-\mathbf{T}_\mathcal{N}) = S_{\mu\nu}^{\mathcal{N}\ominus\mathcal{A}} \tag{11.9}$$

In equation (11.8) the total electronic population of WF i is normalized to one:

$$\sum_{A=1}^{P}\sum_{\mathcal{A}=0}^{L-1} Q_{A\mathcal{A},i} = 1 \qquad \forall i \tag{11.10}$$

where P is the number of atoms in the primitive cell. The information contained in the atomic populations can be employed to characterize the spatial distribution of the i-th WF electrons, by defining the atomic delocalization index λ_i as

$$\lambda_i = \left[\sum_{A=1}^{P}\sum_{\mathcal{A}=0}^{L-1}(Q_{A\mathcal{A},i})^2\right]^{-1} \tag{11.11}$$

The parameter λ_i provides an estimate of the mean number of atoms contributing to WF i [19]. Since WFs are localized functions, for the i-th WF a centroid $\langle\mathbf{r}\rangle_i$ can be found whose spatial coordinates are:

$$\langle r_\alpha\rangle_i = \int d\mathbf{r}|\phi_i(\mathbf{r})|^2 r_\alpha \qquad \alpha = 1,2,3 \tag{11.12}$$

where $\mathbf{r}\equiv(r_1,r_2,r_3)$. The degree of localization of the i-th WF can also be estimated in terms of the corresponding expectation values of the second order moment tensor,

$$\tau_{\alpha\beta}^i = \langle\phi_i|(r_\alpha-\langle r_\alpha\rangle_i)(r_\beta-\langle r_\beta\rangle_i)|\phi_i\rangle \tag{11.13}$$

The trace of $\tau_{\alpha\beta}^i$, $\sigma_i^2 = \sum_{\alpha=1}^3 \tau_{\alpha\alpha}^i$, gives a measure of the spatial spread of WF $\phi_i(\mathbf{r})$.

11.1.2 Local description of the virtual manifold: Projected Atomic Orbitals

Projected Atomic Orbitals (PAOs) are used to span the virtual manifold (according to Pulay's proposal [7]) as WFs are used to span the occupied manifold. Let us define the projector operator onto the occupied HF space $\hat{P}$ as follows:

$$\hat{P} \;=\; \sum_{k}^{occ}|k\,\rangle\langle\,k| = \sum_{\mu\mathcal{M}\nu\mathcal{N}}|\chi_{\mu\mathcal{M}}\,\rangle D_{\mu\mathcal{M}\nu\mathcal{N}}\langle\,\chi_{\nu\mathcal{N}}| \tag{11.14}$$

In the first equality of the above equation the projector $\hat{P}$ is expressed in terms of canonical occupied crystalline orbitals $|k\,\rangle$ while in the second one, via a Brillouin zone integration over all

$|k\rangle$, the projector is expressed in terms of AOs where $D_{\mu\mathcal{M}\nu\mathcal{N}}$ is the density matrix in an AO representation.

The corresponding projector onto the virtual space is then easily obtained: $\hat{Q} = \hat{1} - \hat{P}$. By acting with $\hat{Q}$ on any one of the local functions $\chi_{\mu\mathcal{M}}(\mathbf{r})$ of the original AO set, a Projected AO (or simply PAO) is obtained:

$$\hat{Q}\chi_{\mu\mathcal{M}}(\mathbf{r}) = (\hat{1} - \hat{P})\chi_{\mu\mathcal{M}}(\mathbf{r}) = \tilde{\chi}_{\mu\mathcal{M}}(\mathbf{r}) \tag{11.15}$$

The set of PAOs so generated constitutes a nonorthogonal, linearly dependent, incomplete set of local functions, strictly orthogonal to all WFs. At the same time PAOs are appreciably well localized and keep the symmetry of the parent AOs.

Once a basis set has been selected, to each WF pair is assigned a set of PAOs, those belonging to the corresponding pair domain. Since PAOs are non-orthogonal and usually very diffuse functions, quasi-linear dependencies may take place within that set. To get rid of these redundant functions, the corresponding overlap matrix is diagonalized, and those eigenvectors discarded whose eigenvalues σ are less than a given threshold t_σ that can be set via the **LONTOL** keyword discussed at page 23; the default value is 10^{-4}. The Local OrthoNormal (LON) set of virtual functions is obtained.

The problem of the adequacy of the standard PAO set to describe excitations is a more delicate issue, because the basis sets usually adopted to solve the HF equations for crystals are calibrated so as to describe accurately only the occupied subspace.

11.1.3 Symmetry properties of WFs and PAOs

Symmetry is systematically exploited in all ab initio computational treatments of periodic systems, for instance in CRYSTAL. This happens in two ways, essentially. First, reciprocal space techniques can be used for factorizing the eigenvalue problem of representative matrices of totally symmetric operators (like the Fock or overlap matrices) according to the irreducible representations (irrep) of the group; the same techniques can be used for expressing convolutions over direct lattice vectors as simple products in **k**-space. In both cases, point group symmetry allows us to restrict the problem to the irreducible wedge of the Brillouin zone and, at special **k** points, to further factorize the calculation according to the irreps of that point. Advantage is taken of these opportunities in CRYSCOR for an efficient generalization to periodic systems of the DF technique, as will be shown in Section 11.2.4. Secondly, *symmetry equivalences* exist between various kinds of local quantities which enter the calculation, like products of functions involved in one- and two-electron integrals, which allows us to restrict their estimate to an irreducible subset. The exploitation of the latter type of symmetries is less standard, and plays a central role in the context of local correlation approaches; therefore, it is presented below in some detail. We are interested here in the symmetry properties of our basis functions (AOs, PAOs, WFs), and of their product combinations. Reference is made to the theory of site symmetry in crystals, and to the definition and properties of symmetry adapted Wannier functions (SAWF). The systems considered are insulators that is, their HF manifold consists of fully occupied energy bands.

We shall consider our overcomplete basis set (AOs, PAOs, SAWFs) as composed of "symmetrical objects" (SO), $\boldsymbol{S}_\Lambda$, each associated to a "site" $\boldsymbol{r}_\Lambda$. An SO is an orthonormal set (in our

jargon, a "flower)" of n_Λ local functions (its "petals"), which are a basis of an n_Λ-dimensional representation α_Λ of the point group G_Λ of site $\boldsymbol{r}_\Lambda$ (site group). The group G_Λ is a subgroup of the point group of the crystal. For example, a shell of ℓ-type AOs on a given atom forms an SO of dimensionality $n_\Lambda = 2\,\ell + 1$, having its site in the nuclear position. An operator of G_Λ acting on the petals of $\boldsymbol{S}_\Lambda$ will simply effect a unitary transformation among the latter (that is, the SO is left unchanged). An operator not belonging to G_Λ (in particular, a pure translation) will instead transform it into another flower, equivalent to the previous one, apart (possibly) from a translation and a unitary transformation among its petals. All the flowers so generated from a reference one with its site in the zero cell, $\boldsymbol{S}_{\Lambda,0}$, will form a "bunch" , $\overline{\Lambda}$, which contains all images of the parent flower in the other cells, $\boldsymbol{S}_{\Lambda,\mathcal{G}}$, but may also comprise more than one flower per cell. The bunch is a basis of an induced representation (indrep) of the space group.
The effect of the general space group operator $\widehat{V}$ onto an SO can then be formulated as follows:

$$\widehat{V}\ \boldsymbol{S}_{\Lambda,\mathcal{G}}\ =\ \boldsymbol{W}_V^\Lambda\ \boldsymbol{S}_{\Lambda_V,\mathcal{G}_V^\Lambda} \tag{11.16}$$

The cell, $\mathcal{G}_V^\Lambda$, the label Λ_V, identifying the specific SO within the cell, and the unitary $\boldsymbol{W}_V^\Lambda$ matrix, corresponding to the representation α_Λ of dimension n_Λ, are determined by the program following a group theoretical analysis. Vice versa, given any two flowers of the bunch, there exists at least one $\widehat{V}$ which carries the first into the second one.
If the AOs in the basis set are centered in the crystal atoms, equivalent atoms having equivalent sets, then each AO shell constitutes an SO, as anticipated. Since the $\widehat{Q}$ projector has the full symmetry of the crystal ($\forall\ \widehat{V}\ :\ [\widehat{Q},\widehat{V}] = 0$), shells of PAOs have the same symmetry as the parent AO shells. The standard basis set of GTFs used in Cryscor, is therefore perfectly suited for the present purposes.
The characterization of WFs as SOs is less obvious; appropriate SAWFs have to satisfy the following conditions.

1. The number of SAWFs per cell must equal the number of filled bands explicitly treated; the set of all SAWFs must span the same manifold as the states belonging to those bands.

2. They must form an orthonormal set (which is not required for PAOs).

3. They must be classifiable as SOs, in the sense that they transform according to equation (11.16).

4. Among different SAWF sets satisfying the above conditions, the one should be preferred which maximizes the number of symmetry equivalences between those quantities which are most important in determining the computational cost (see below).

5. SAWFs should be as local as possible owing to some localization criterion.

Satisfying all these conditions is not trivial, and the solution is often not unique. A strategy which has proved usually effective has been proposed where the construction of the SAWFs is guided by the localization procedure [6]. Localized functions are classified as bond, atomic, etc, relying on their atomic populations, and symmetrized accordingly. However, the functional adopted

(Ruedenberg, Boys [18], Pipek-Mezey [19], etc.) may be influential on the type of symmetry of the SAWF set which is obtained. The following can be noted for this purpose.
At variance with PAOs, the sites of the SAWF flowers may not coincide with atomic nuclei (the case of diamond is an example). In particular, if the site is at a general point, the bunch of that SAWF contains as many single-petal flowers per cell as is the order of the point group of the crystal. Considering all possible sites and their indreps may result in some cases in a large number of *types* of SO.
A variety of combinations of such SOs can be selected, in principle, compatible with condition 1 in the list above. Among them, the best one should be preferred owing to criterion 4, and a trial set of orthonormal SAWFs is constructed accordingly. Finally, unitary transformations should be performed among them so as to maximize localization while preserving the symmetry characterization.

11.1.4 Wannier functions: distances and domains

As will be clarified in the following, the main approximation of the LMP2 method (or more generally of every method formulated within a local approach) is the truncation of various sums, running in principle over the whole (infinite) sets of WFs and PAOs of the crystalline system, in the fundamental equations. The sum over WFs can be truncated according to a distance criterion while for each pair of WFs the sum over PAOs can be limited to virtual functions lying in a spatial vicinity of the two WFs of the pair. As a first step, it is then necessary to introduce for any pair $(\phi_{i\mathcal{I}}(\mathbf{r}),\ \phi_{j\mathcal{J}}(\mathbf{r}))$ of WFs a measure $D_{i\mathcal{I},j\mathcal{J}}$ of the *distance* between them.

A *kernel* $[i0]^0$ can be associated to each reference WF $\phi_{i0}(\mathbf{r})$ and then transferred to the whole set $\{\phi_i(\mathbf{r})\}$ of its periodic images in the crystal. This minimum domain is defined as the set of any atom of the system contributing to the total Mulliken electronic population of the WF $\phi_{i0}(\mathbf{r})$ for more than a given threshold Q^0. In other words, remembering equation (11.8), only atoms $A\mathcal{A}$ satisfying the following equation are included in the *kernel* $[i0]^0$:

$$[i0]^0 = \{A\mathcal{A}\} \quad \backslash \quad Q_{A\mathcal{A},i0} = \sum_{\mu\in A\mathcal{A}} \sum_{\nu=1}^{M} \sum_{\mathcal{N}=0}^{L-1} L^{\mathcal{A}}_{\mu,i} \left(L^{\mathcal{N}}_{\nu,i}\right)^* S_{\mu\mathcal{A}\nu\mathcal{N}} \geq Q^0 \tag{11.17}$$

In the Cryscor program Q^0 has the default value 0.1 but it can be modified by the user with the **MINPOP** keyword (see page 17). If Q^0 is reasonably chosen then the *kernel* represents a minimal chemical description of the corresponding WF. If the selected WF is a *bond* one then the corresponding *kernel* will be the union of the two atoms belonging to the bond; otherwise if the selected WF is an *atomic* one, then the corresponding *kernel* will contain only one atom (the one on which the WF is centred). The distance $D_{i\mathcal{I},j\mathcal{J}}$ between two general WFs can be defined as the minimum distance between any two atoms belonging to the respective *kernels* $[i\mathcal{I}]^0 = \{B\mathcal{B}\}$ and $[j\mathcal{J}]^0 = \{B'\mathcal{B}'\}$:

$$D_{i\mathcal{I},j\mathcal{J}} = \min_{\{B\mathcal{B}\}\{B'\mathcal{B}'\}} |\mathbf{s}_B - \mathbf{s}_{B'} + \mathbf{T}_{\mathcal{B}} - \mathbf{T}_{\mathcal{B}'}| \tag{11.18}$$

It follows that a classification of WFs pairs by distance is now possible. Since excitations from two local occupied orbitals (as suggested from molecular quantum chemical experience) become

progressively less important with increasing distance between the two centers, the classification of WFs according to their relative distance leads to the possible exploitation of a hierarchical treatment of WFs pairs (and of the corresponding integrals which constitute the real computational bottleneck of the method).

Wannier functions pairs can be classified into *strong*, *weak*, *distant* and *very distant* pairs according to the following scheme:

$$\begin{cases} D_{i\mathcal{I},j\mathcal{J}} = 0 & \textit{strong} \\ 0 < D_{i\mathcal{I},j\mathcal{J}} \leq d_1 & \textit{weak} \\ d_1 < D_{i\mathcal{I},j\mathcal{J}} \leq d_2 & \textit{distant} \\ D_{i\mathcal{I},j\mathcal{J}} > d_2 & \textit{very distant} \end{cases}$$

where d_1 and d_2 are two parameters to be carefully tuned. Our default values are $d_1 = 8$ Å and $d_2 = 12$ Å. For most purposes, strong and weak pairs are treated in CRYSCOR as a unique set, that of *close-by* pairs. The 2-electron repulsion integrals (ERI) corresponding to *close-by* pairs are either exactly solved or accurately and ecomomically evaluated via a density fitting periodic (DFP) scheme [11, 12], which essentially consists in expressing product distributions as a linear combination of auxiliary fitting functions. ERIs associated to *distant* pairs are treated via a multipolar approximation, which can safely be used only in the case of almost zero overlap between the interacting product distributions. Finally, the *very distant* pairs are totally excluded from the explicit computation; this condition is crucial for the asymptotically linear scaling of computational resources versus the size of the system. However, the contribution to the correlation energy from the neglected pairs can be recovered by exploiting the long range r^{-6} dependence of pair-energies with respect to their distance, via a Lennard-Jones extrapolation technique (see the keyword **LENJONES** at page 18 and Section 11.2.3).

As mentioned in the previous pages it is mandatory to associate an excitation domain $[i0]$ to each *reference* WF $\phi_{i0}(\mathbf{r})$. The general i-th domain consists of a set of atoms $\{A\mathcal{A}\}_i$: when atom $A\mathcal{A}$ is said to belong to a given domain it is intended that the full set of PAOs centred on atom A (in cell $\mathcal{A}$) belongs to that domain. As it happens for *kernels*, also excitation domains are transferred to the whole set $\{\phi_i(\mathbf{r})\}$ of periodic images of $\phi_{i0}(\mathbf{r})$ in the crystal. In the CRYSCOR program there are different ways to define WF domains.

An explicit definition is feasible via the **DOMDEF** keyword (see page 21), by specifying for each WF the set of atoms $\{A\mathcal{A}\}_i$. This method is user-friendly for systems with few symmetry unique WFs, otherwise it is cumbersome and may possibly lead to mistakes.

When we are dealing with molecular crystals (with m molecules M_1, M_2, ..., M_m in the reference cell) it can be convenient to work with molecular domains (**DOMMOL** keyword, see page 22) that is to say, to include in each domain of a WF belonging to the j-th molecule M_j the whole set of atoms of that molecule.

Still another criterion can be used, based on the one proposed in a molecular context by Boughton and Pulay [22] and explained in detail below: it can be activated using the **DOMPUL**

keyword (see page 21).

Boughton-Pulay automatic excitation domain selection

For a given WF, some atoms are selected whose combined basis functions best span the WF subspace. The spanning criterion is based upon a least-squares fit of the functions on the selected atoms to the WF. A selection of atoms (of functions) is required to span its corresponding WF by at least 98%, judged by the least-squares residual [22]. In the CRYSCOR program the default value of this spanning parameter T^{BP} is then 0.98 but it can be changed via a dedicated input card.

For a given *reference* WF $\phi_{i0}(\mathbf{r})$, the least-squares residual (which has to be less than $1-T^{BP}$) for a *proposed selection* Σ_{i0} of atoms (and consequently of PAOs) is

$$f(\Sigma_{i0}) = \min\left[\int d\mathbf{r}\,(\phi_{i0}(\mathbf{r}) - \phi'_{i0}(\mathbf{r}))^2\right] \tag{11.19}$$

where

$$\begin{aligned}\phi_{i0}(\mathbf{r}) &= \sum_{\mu\mathcal{M}} L^{\mathcal{M}}_{\mu,i0}\chi_\mu(\mathbf{r}-\mathbf{s}_\mu-\mathbf{T}_\mathcal{M}) \\ \phi'_{i0}(\mathbf{r}) &= \sum_{\nu\mathcal{N}\in\Sigma_{i0}} \sigma^{\mathcal{N}}_{\nu,i0}\chi_\nu(\mathbf{r}-\mathbf{s}_\nu-\mathbf{T}_\mathcal{N})\end{aligned} \tag{11.20}$$

For this purpose we must solve the following set of equations for all $\sigma^{\mathcal{B}}_{\beta,i0}$:

$$\frac{\partial f(\Sigma_{i0})}{\partial \sigma^{\mathcal{N}}_{\nu,i}} = -2\sum_{\mu\mathcal{M}} L^{\mathcal{M}}_{\mu,i0} S_{\mu\mathcal{M}\nu\mathcal{N}} + 2\sum_{\beta\mathcal{B}\in\Sigma_{i0}} \sigma^{\mathcal{B}}_{\beta,i0} S_{\nu\mathcal{N}\beta\mathcal{B}} = 0 \tag{11.21}$$

or in matrix notation

$$\mathbf{S}^{\Sigma_{i0}}\boldsymbol{\sigma}_{i0} = \mathbf{v}_{i0} = \mathbf{S}^{\diamond}\mathbf{L}_{i0} \tag{11.22}$$

where $\mathbf{S}^{\Sigma_{i0}}$ is the overlap matrix of the collected elements whose rows and columns span the functional space Σ_{i0}, $\mathbf{L}_{i0}$ and $\boldsymbol{\sigma}_{i0}$ are the complete and partial WF coefficients matrices respectively, $\mathbf{S}^{\diamond}$ is the overlap matrix whose columns span the space Σ_{i0} but whose rows span the full crystalline basis. The least-squares residual of the proposed atomic selection for WF $\phi_i(\mathbf{r})$ is then:

$$f(\Sigma_{i0}) = 1 - \boldsymbol{\sigma}^T_{i0}\mathbf{v}_{i0} = 1 - \mathbf{L}^T_{i0}\mathbf{S}^{\diamond T}\mathbf{S}^{\Sigma^{-1}_{i0}}\mathbf{S}^{\diamond}\mathbf{L}_{i0} \tag{11.23}$$

where we used the usual convention $\mathbf{A}^T_{ij} = \mathbf{A}_{ji}$ for any matrix $\mathbf{A}$. Equation (11.23) can be rewritten more explicitly in the following way:

$$f(\Sigma_{i0}) = 1 - \sum_{\mu\mathcal{M}\in\Sigma_{i0}}\sum_{\nu\mathcal{N}\in\Sigma_{i0}}\sum_{\alpha\mathcal{A}}\sum_{\beta\mathcal{B}} L_{i0,\alpha\mathcal{A}\oplus\mathcal{N}} S^{\diamond}_{\alpha\mathcal{A},\nu 0} S^{\Sigma^{-1}_{i0}}_{\nu\mathcal{N},\mu\mathcal{M}} S^{\diamond}_{\mu 0,\beta\mathcal{B}} L_{\beta\mathcal{B}\oplus\mathcal{M},i0} \tag{11.24}$$

11.2 Local-MP2 method for periodic systems

The following notations will be adopted:

1. Standard Latin or Greek indices (μ, i, a, P, etc.) label functions within a given unit cell.
2. Calligraphic indices $\mathcal{M}$, $\mathcal{I}$, $\mathcal{A}$, $\mathcal{P}$ denote lattice vectors and the corresponding crystal cell.
3. Bold Latin or Greek indices combine function and cell index: $\mathbf{i} \equiv (i\mathcal{I})$, $\mathbf{a} \equiv (a\mathcal{A})$, etc.
4. $\chi_{\mu\mathcal{M}}$, $\chi_{\nu\mathcal{N}}$, ... or equivalently $\mu\mathcal{M}$, $\nu\mathcal{N}$, ... are atomic orbitals (AOs).
5. $\psi_{a\mathcal{A}}$, $\psi_{b\mathcal{B}}$, ... or equivalently $a\mathcal{A}$, $b\mathcal{B}$, ..., or $\mathbf{a}$, $\mathbf{b}$, ... are projected atomic orbitals (PAOs).
6. $\phi_{i\mathcal{I}}$, $\phi_{j\mathcal{J}}$, ... or equivalently $i\mathcal{I}$, $j\mathcal{J}$, ..., or $\mathbf{i}$, $\mathbf{j}$, ... are Wannier functions (WFs).
7. $\rho_{i\,a\mathcal{A}} = \phi_{i0}\phi_{a\mathcal{A}}$ is the WF-PAO product density.
8. $\Xi_{P\mathcal{P}}$, $\Xi_{Q\mathcal{Q}}$, ... or equivalently $P\mathcal{P}$, $Q\mathcal{Q}$ are fitting functions.

11.2.1 Local MP2 Energy

Let us recall here the expression of the MP2 energy as obtained within the canonical approach:

$$\begin{aligned} E^{MP2} &= \sum_{i}^{occ.}\sum_{j>i}^{occ.}\sum_{a}^{vir.}\sum_{b>a}^{vir.} \frac{\langle \Psi^{HF}|H|\Psi_{ij}^{ab}\rangle\langle\Psi_{ij}^{ab}|H|\Psi^{HF}\rangle}{E^0 - E^0\binom{ab}{ij}} = \\ &= \sum_{i}^{occ.}\sum_{j>i}^{occ.}\sum_{a}^{vir.}\sum_{b>a}^{vir.} \frac{|\langle ij||ab\rangle|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \end{aligned} \tag{11.25}$$

In our LMP2 method, the local functions representing occupied and virtual manifolds of the HF reference are constructed differently. The occupied space is spanned by orthonormal WFs, generated from the canonical Bloch orbitals produced by CRYSTAL using the Localization-Wannierization scheme by Zicovich-Wilson *et. al* [5]. The WFs are symmetrized *a posteriori* according to the procedure suggested by Casassa *et al.* [6] For the virtual manifold, as suggested by Pulay [7], we use PAOs, which are AOs projected out from the occupied space. PAOs form a redundant nonorthogonal set, but at the same time are appreciably well localized and have the symmetry of the parent AOs.

The orbital invariant MP2 energy per unit cell in the local approximation is given by the following expression:

$$E^{\mathrm{LMP2}} = \sum_{i}\sum_{\{j\mathcal{J}\,\mathrm{near}\,i\}} E_{i,j\mathcal{J}}^{\mathrm{LMP2}} \tag{11.26}$$

$$E_{i,j\mathcal{J}}^{\mathrm{LMP2}} = \sum_{(a\mathcal{A},b\mathcal{B})\in[i,j\mathcal{J}]} K_{a\mathcal{A},b\mathcal{B}}^{i,j\mathcal{J}} \left(2\,T_{a\mathcal{A},b\mathcal{B}}^{i,j\mathcal{J}} - T_{b\mathcal{B},a\mathcal{A}}^{i,j\mathcal{J}}\right) \tag{11.27}$$

Here $K^{i,j\mathcal{J}}_{a\mathcal{A},b\mathcal{B}} \equiv (i0, a\mathcal{A} \,|\, j\mathcal{J}, b\mathcal{B})$ are the ERIs, and $T^{i,j\mathcal{J}}_{a\mathcal{A},b\mathcal{B}}$ are the excitation amplitudes. The only approximation introduced is the truncation of the two sums in these equations.
Consider first equation (11.27). For a given pair of WFs **i** and **j**, the excitation space is confined to their spatial vicinity or, in other words, to the so called PAO domain [**ij**] of that pair. This means that the amplitudes $T^{\mathbf{ij}}_{\mathbf{ab}}$ are assumed to be non-zero only if the PAOs **a** and **b** are spatially close to the WFs **i** or **j**. To validate this approximation we note that the integrals $K^{\mathbf{ij}}_{\mathbf{ab}}$ and amplitudes $T^{\mathbf{ij}}_{\mathbf{ab}}$ both decay exponentially with the distance between, say, **i** and **a**. The former follows from the exponential decay of the WFs and PAOs, and the latter from the decay of the corresponding integrals, as is clearly seen from the orbital invariant MP2 equations (see Section 11.2.2). The energy expression (11.27) squares this decay. There are different ways to determine the PAO domains. One usually employs the Boughton-Pulay procedure, which chooses the atoms to be included in the domain (and all the corresponding AOs) so that the orbital populations on these atoms sum up to a certain threshold.
WF **i** in the energy expression (11.26) is located in the reference cell, while the range of the WFs **j** in principle spans the whole space. Since WFs and PAOs are mutually orthogonal, the decay of the integrals $K^{\mathbf{ij}}_{\mathbf{ab}}$ and of the amplitudes $T^{\mathbf{ij}}_{\mathbf{ab}}$ with respect to the distance between the WFs **i** and **j** is r^{-3}. However, in the energy expression (11.27) this decay is again squared and so the energy contributions decay according to the well-known London's law as r^{-6}. As a consequence, for each **i** the list of **j**'s can be restricted to a set {**j** near **i**} beyond which the energy contributions are negligibly small. In molecular LMP2 applications the cut-off radius (i.e., the maximum distance between **i** and **j**) is usually set to 15 Bohr. In 2D and especially 3D crystals, the situation is somewhat different, because of their compactness. It will be shown in Section 11.2.3 how to evaluate all contributions up to infinity, while restricting the actual computation to WF pairs within a distance of 15-20 Bohr.

11.2.2 Local MP2 amplitudes

In order to obtain the LMP2 energy according to (11.26) and (11.27), one needs to evaluate the amplitudes $T^{\mathbf{ij}}_{\mathbf{ab}}$ and the integrals $K^{\mathbf{ij}}_{\mathbf{ab}}$. We postpone the discussion of the approaches to the integral calculation to section 11.2.4, and focus here on the procedure of evaluation of the LMP2 amplitudes for a precomputed set of the integrals.

The MP2 orbital-invariant amplitude equations in an orthogonal basis take the form:

$$0 \;=\; R^{\mathbf{ij}}_{\mathbf{ab}} = K^{\mathbf{ij}}_{\mathbf{ab}} + \sum_{\mathbf{c}} (F_{\mathbf{ac}} T^{\mathbf{ij}}_{\mathbf{cb}} + T^{\mathbf{ij}}_{\mathbf{ac}} F_{\mathbf{cb}}) - \sum_{\mathbf{k}} (F_{\mathbf{ik}} T^{\mathbf{kj}}_{\mathbf{ab}} + T^{\mathbf{ik}}_{\mathbf{ab}} F_{\mathbf{kj}}). \tag{11.28}$$

Here $R^{\mathbf{ij}}_{\mathbf{ab}}$ are the residuals, which become zero at convergence, and $F_{\mathbf{ab}}$ and $F_{\mathbf{ij}}$ the Fock matrix elements in the PAO and WF subspaces. By solving this system of linear equations one obtains the MP2 amplitudes. From (11.28) it clearly follows that at convergence, owing to the exponential decay of the Fock matrix, the decay rate of the amplitudes is dictated by that of the integrals, as was mentioned in the previous subsection. When the local approximations on the pair list {**ij**} and orbital domains [**ij**] are invoked, the number of unknowns $T^{\mathbf{ij}}_{\mathbf{ab}}$ becomes smaller than the number of equations. In order to solve the system, the residuals (and the corresponding equations) which

do not map their amplitude counterparts are disregarded. This restriction actually is equivalent to formulating the MP2 theory within the set of doubly excited determinants, truncated accordingly to the local approximations introduced.
For non-orthogonal virtual orbitals (PAOs), the LMP2 equations take a more complicated form:

$$R^{\mathbf{ij}}_{\mathbf{ab}} \;=\; K^{\mathbf{ij}}_{\mathbf{ab}} + A^{\mathbf{ij}}_{\mathbf{ab}} + B^{\mathbf{ij}}_{\mathbf{ab}} \;, \tag{11.29}$$

with the matrices $A^{\mathbf{ij}}_{\mathbf{ab}}$ and $B^{\mathbf{ij}}_{\mathbf{ab}}$ defined as follows:

$$A^{i,j\mathcal{J}}_{a\mathcal{A},b\mathcal{B}} \;=\; \sum_{(c\mathcal{C},d\mathcal{D})\in[i,j\mathcal{J}]} \Big[F^{\mathcal{C}\ominus\mathcal{A}}_{ac}\, T^{i,j\mathcal{J}}_{c\mathcal{C},d\mathcal{D}}\, S^{\mathcal{B}\ominus\mathcal{D}}_{db} + S^{\mathcal{C}\ominus\mathcal{A}}_{ac}\, T^{i,j\mathcal{J}}_{c\mathcal{C},d\mathcal{D}}\, F^{\mathcal{B}\ominus\mathcal{D}}_{db} \Big] \tag{11.30}$$

$$B^{i,j\mathcal{J}}_{a\mathcal{A},b\mathcal{B}} \;=\; -\sum_{(c\mathcal{C},d\mathcal{D})\in u[i]} S^{\mathcal{C}\ominus\mathcal{A}}_{ac}\, \beta'^{\,i,j\mathcal{J}}_{c\mathcal{C},d\mathcal{D}}\, S^{\mathcal{B}\ominus\mathcal{D}}_{db} - \sum_{(c\mathcal{C},d\mathcal{D})\in u[j\mathcal{J}]} S^{\mathcal{C}\ominus\mathcal{A}}_{ac}\, \beta''^{\,i,j\mathcal{J}}_{c\mathcal{C},d\mathcal{D}}\, S^{\mathcal{B}\ominus\mathcal{D}}_{db} \tag{11.31}$$

$$\beta'^{\,i,j\mathcal{J}}_{c\mathcal{C},d\mathcal{D}} = \sum_{\{k\mathcal{K}\,\mathrm{near}\,i\}} T^{i,k\mathcal{K}}_{c\mathcal{C},d\mathcal{D}}\, F^{\mathcal{J}\ominus\mathcal{K}}_{kj} \tag{11.32}$$

$$\beta''^{\,i,j\mathcal{J}}_{c\mathcal{C},d\mathcal{D}} = \sum_{\{k\mathcal{K}\mathrm{near}\,j\mathcal{J}\}} F^{\mathcal{K}}_{ik}\, T^{k\mathcal{K},j\mathcal{J}}_{c\mathcal{C},d\mathcal{D}} \tag{11.33}$$

Here $S^{\mathcal{G}}_{ab}$ ($F^{\mathcal{G}}_{ab}$) denotes the element of the overlap (Fock) matrix between functions $a0$ and $b\mathcal{G}$, and the simplifications due to translational symmetry are implicitly introduced. The sum in equation (11.30) runs over all PAOs in the WF pair domain $[i, j\mathcal{J}]$. In equation (11.31) the sums are running over the "united domains" $u[i]$ or $u[j\mathcal{J}]$:

$$u[i] = \bigcup_{k\mathcal{K}} [i0\, k\mathcal{K}] \quad , \quad u[j\mathcal{J}] = \bigcup_{k\mathcal{K}} [j\mathcal{J}\, k\mathcal{K}] \;, \tag{11.34}$$

with the union comprising not only symmetry-unique pairs WFs (see Section 11.1.3). Finally, the sums in equations (11.32) and (11.33) run over all WFs which can form pairs with i or $j\mathcal{J}$, respectively.
The cost for computing the B-term of the residual (equation 11.31) scales as $O(N^2)$, where N is the number of pairs. In 3D crystals this corresponds to R^6_{cut} scaling. (R_{cut} is the cutoff distance for the pairs included in the calculation), as the number of pairs increases cubically with the cutoff distance. Moreover, although the number of different $i, j\mathcal{J}$ pairs can be reduced by considering only the N_{irr} symmetry irreducible ones (see section 11.1.3), in the sums of equations (11.32) and (11.33) one has to take into account *all* possible WFs $k\mathcal{K}$ forming pairs with i or $j\mathcal{J}$. Advantage can be taken, however, of the sparsity of the f, S, T matrices. Exploiting this sparsity leads to a reduction of the overall scaling from $O(N_{\mathrm{irr}}N)$ to just $O(N_{\mathrm{irr}})$.

We solve the system of the equations using the Gauss-Seidel iterative scheme. The procedure is similar to one used in the molecular local methods. Briefly, it consists of the following:

1. Evaluation of the residual matrix for a given pair (**ij**).

2. Transformation of the residual matrix into the so-called pseudo-canonical basis set, an orthonormal set of virtuals which spans the space of the domain [**ij**], and diagonalizes the Fock matrix in this space.

3. Calculation of the updates for the amplitudes of the pair (**ij**) (in the pseudo-canonical set the expression for the updates takes a particularly simple form).

4. Back-transformation of the updates to the PAO basis.

In step 1, the most recently updated amplitudes of other pairs (B-terms, equations (11.32), (11.33)) are used.
It follows from equation (11.31) that evaluating the residuals for a given pair involves also amplitudes corresponding to many other pairs. In order to avoid excessive input/output, we keep all the amplitudes in memory. The size of the amplitude vector depends quadratically on the size of the domains, and therefore when extended domains or high-quality orbital basis sets are used, such storage may become a memory bottleneck. In this case one would have to save amplitudes on disk and invoke a paging algorithm.
The computational complexity of the matrix multiplications in the evaluation of the residuals scales cubically with respect to the domain size. So again, for large domains or basis sets, the computational time needed for solving the LMP2 equations can become dominant. There is an interesting technique of Laplace transformed LMP2 formalism, which provides the amplitudes without solving the LMP2 equations (a Laplace quadrature for the amplitudes must be evaluated numerically instead). Interestingly, this approach allows the restriction on the residuals to be released, which gives usually an energy closer to the canonical MP2 one. An adaptation of this scheme to the periodic case is a matter of future work.

High quality correlation calculations require the use of rather extended basis sets (at least of triple-zeta quality and for weakly bound systems also with diffuse functions). However, such basis sets can cause serious difficulties when solving the HF equations for periodic systems, due to quasi-linear-dependence and convergence problems. In order to satisfy these contradictory requirements we have introduced the dual basis scheme, where different basis sets for HF and LMP2 can be selected. In such case, however, the Brillouin theorem is no longer valid, since the LMP2 reference is not actually the HF reference for the LMP2 basis set. Hence non-zero singles amplitudes $T^{\mathbf{i}}_{\mathbf{a}}$ also contribute to the energy:

$$E^{\mathrm{LMP2}}_{\mathrm{singles}} = \sum_{i,a\mathcal{A}} F_{i,a\mathcal{A}}\, T^{i}_{a\mathcal{A}} \tag{11.35}$$

If the HF basis set is of a reasonable quality (say double-zeta), the occupied-virtual blocks of the Fock matrix are small. Then the equations for the singles amplitudes take the form:

$$0 = R^{\mathbf{i}}_{\mathbf{a}} = F_{\mathbf{ia}} + \sum_{\mathbf{b}} F_{\mathbf{ab}}\, T^{\mathbf{i}}_{\mathbf{b}} - \sum_{\mathbf{b\,j}} F_{\mathbf{ij}}\, T^{\mathbf{j}}_{\mathbf{b}}\, S_{\mathbf{ba}}. \tag{11.36}$$

The equations for the singles and doubles amplitudes are completely decoupled, and the singles part is computationally cheap.

11.2.3 Lennard-Jones extrapolation for very distant pairs

One of the distinctive features of local correlation schemes in molecular calculations is that one can safely ignore the contributions E_{ij} from pairs which are "very distant" from each other, that

is, when the distance between the respective core domains exceeds a pre-fixed value d_2 (typically, $d_2 = 15$ Bohr). This approximation is crucial to warrant N-scaling of computational costs, and is justified by the fact that at large distances pair energies follow the London D_{ij}^{-6} law as is present in the well known 6-12 Lennard-Jones expression, D_{ij} being the distance between the centers of the two distributions. However, in 3D crystals the number of pairs grows quadratically with the distance d and the contribution to the correlation energy from all pairs at a distance $D_{ij} \geq d$ falls off merely as d^{-3}. Considerably larger cutoff distances d_2 should be adopted with respect to molecular calculations in order to achieve comparable accuracy, with a formidable impact on computational costs. This problem is easily circumvented by taking advantage of the simple law of inter-pair decay rate and of translational symmetry so as to extrapolate to infinity the sum of far-off contributions. The procedure adopted by us is as follows.

1. Each symmetry irreducible pair of SAWF flowers, $(I0, J\mathcal{J})$, is characterized by a distance $D_{I,J\mathcal{J}} = |\mathbf{r}_{I0} - \mathbf{r}_{J0} - \mathbf{G}|$, where $\mathbf{r}_{X0}$ are the sites of the respective SAWFs in the zero cell (see section 11.1.3: note that this definition of distance between *WF flowers* is different from that adopted in equation 11.18 for the classification of WF pairs according to proximity). Its contribution to the correlation energy, $E_{I,J\mathcal{J}}$, is the sum of all the terms as in equation 11.27 between the respective petals

2. After performing an MP2 calculation within a distance d_2, the different irreducible flower pairs are classified into separate sets $M(\overline{IJ})$ according to the pair of bunches $\overline{I}$, $\overline{J}$ the two flowers belong to, irrespective of the order.

3. Within each $M(\overline{IJ})$ set, only those pairs are considered for which $d^{lj} \leq D_{I,J\mathcal{J}} \leq d_2$, where d^{lj} is a distance set from input such that the asymptotic London d^{-6} behaviour is deemed to be approximately satisfied beyond it (typically, d^{lj}=10 Bohr). If there are enough pairs in the set, a best fit Lennard-Jones coefficient $C^{lj}_{(\overline{IJ})}$ is determined, by minimizing the quantity:

$$S = \sum_{(I,J\mathcal{J}) \in M(\overline{IJ})} \left[E_{I,J\mathcal{J}} - \widetilde{E}^{lj}_{(\overline{IJ})}(D_{I,J\mathcal{J}}) \right]^2 \tag{11.37}$$

where:

$$\widetilde{E}^{lj}_{(\overline{IJ})}(d) = C^{lj}_{(\overline{IJ})}\, d^{-6} \tag{11.38}$$

Note that in a general case, for two pairs in the same set and at the same distance, the energy may not coincide because the two flowers have different respective orientations; the coefficient determined is then an average between all possible orientations.

4. A radius $D > d_2$ is set from input. The contribution from all flower pairs within D, but *not* explicitly included in the calculation is evaluated by direct summation using expression (11.38) with the appropriate coefficients. The contribution $E^{\text{res}}(D)$ from all residual pairs is finally obtained by integration up to infinite distance:

$$E^{\text{res}}(D) = \frac{4\,\pi}{3\,V_{\text{cell}}} \left[\sum_{\overline{I},\overline{J}} C^{lj}_{(\overline{IJ})}\, n_{\overline{I}}\, n_{\overline{J}} \right] D^{-3}\,, \tag{11.39}$$

where V_{cell} is the volume of the unit cell of the crystal, and $n_{\overline{I}}$, $n_{\overline{J}}$ are the number of flowers per cell of the respective bunches.

5. A similar formula is used for 2-dimensional structures.

This scheme can be activated by the **LENJONES** keyword.

11.2.4 The (ia | jb) integrals

In this section we discuss one of the most essential issues of the LMP2 method, the evaluation of the ERIs (**ia** | **jb**), which are occurring in both the energy and amplitude (Section 11.2.2) equations. The techniques employed in CRYSCOR for dealing with this problem differ according to the distance between the two WFs, **i** and **j**. For *close-by* pairs ($D_{ij} \leq d_1$) a Density Fitting Periodic technique (DFP), as described in Section 11.2.4, is the recommended option, although by default ERIs are evaluated analytically (which can serve for calibration purposes).

At intermediate distances (distant pairs), the multipolar approximation provides a very efficient and accurate solution of the problem (see Section 11.2.4). Finally, for very distant pairs where pair energies follow the r^{-6} regime, the ERIs are not needed because their contribution to energy can be evaluated by extrapolation techniques, as shown in Section 11.2.3.

Density Fitting Periodic (DFP) technique for strong and weak pairs

The applicability of our method to systems of general interest is limited by the computational resources needed for the evaluation of electron repulsion integrals (ERI) between two Product Densities (PDs) of a WF times a PAO. This bottleneck arises from the very nature of WFs and PAOs, which are expressed as a linear combination of the AO basis functions each. It becomes especially severe in the crystalline environment where the functions, although localized, are extended in principle to infinity. A way to overcome this difficulty is offered by exploiting density fitting method for periodic systems. This technique represents an extension of the DF approach implemented in the molecular context to crystals, and allows a reduction of the overall CPU time by up to three orders of magnitude [11, 12, 13].

Two approaches are implemented, which are both stable and have, on average, comparable performance. These are the direct-space fitting and the reciprocal-space fitting, activated by the **DIRECT** and **KSPACE** keywords.

The basic concept is to approximate each PD as a linear combination of auxiliary functions Ξ_P, here and in the following called fitting functions (FF):

$$\begin{aligned} \rho_{ia\mathcal{A}} &\equiv \sum_{\mu\mathcal{M},\nu\mathcal{N}} c_i^{\mu\mathcal{M}}\, c_a^{\nu\mathcal{N}\ominus\mathcal{A}}\, \phi_{\mu\mathcal{M}}\, \phi_{\nu\mathcal{N}} \\ &\approx \sum_{P\mathcal{P}} d_{ia\mathcal{A}}^{P\mathcal{P}}\, \Xi_{P\mathcal{P}} \equiv \widetilde{\rho}_{i\,a\mathcal{A}} \end{aligned} \tag{11.40}$$

where the coefficients $d_{ia\mathcal{A}}^{P\mathcal{P}}$ of the linear combination are determined by minimizing a suitable error functional, expressed as:

$$\begin{aligned} \Delta_{ia\mathcal{A}} &= (\rho_{ia\mathcal{A}} - \widetilde{\rho}_{ia\mathcal{A}} \,|\, w_{12} \,|\, \rho_{ia\mathcal{A}} - \widetilde{\rho}_{ia\mathcal{A}}) = \\ &= (\Delta\rho_{ia\mathcal{A}} \,|\, w_{12} \,|\, \Delta\rho_{ia\mathcal{A}}) \end{aligned} \tag{11.41}$$

Common choices for w_{12} are r_{12}^{-1} or δ_{12}, corresponding to the Coulomb or overlap metric, respectively; here and in the following we will refer to the former, since it leads to a simpler formalism, and is the metric used in the present implementation. The equation determining the coefficients then reads:

$$d_{ia\mathcal{A}}^{P\mathcal{P}} = \sum_{Q\mathcal{Q}} \left[J^{-1}\right]_{P\mathcal{P},Q\mathcal{Q}} \; J_{ia\mathcal{A}}^{Q\mathcal{Q}} \quad , \tag{11.42}$$

where we have introduced the notations:

$$\begin{aligned} J_{P\mathcal{P},Q\mathcal{Q}} &= \int d\mathbf{r}_1 \int d\mathbf{r}_2 \, \frac{\Xi_{P\mathcal{P}}(\mathbf{r}_1)\, \Xi_{Q\mathcal{Q}}(\mathbf{r}_2)}{\mathbf{r}_{12}} \; ; \\ J_{ia\mathcal{A}}^{Q\mathcal{Q}} &= \int d\mathbf{r}_1 \int d\mathbf{r}_2 \, \frac{\rho_{ia\mathcal{A}}(\mathbf{r}_1)\, \Xi_{Q\mathcal{Q}}(\mathbf{r}_2)}{\mathbf{r}_{12}} \end{aligned} \tag{11.43}$$

Each ERI is then approximated as:

$$\begin{aligned} K_{a\mathcal{A},\,b\mathcal{B}}^{i,\,j\mathcal{J}} &\approx \widetilde{K}_{a\mathcal{A},\,b\mathcal{B}}^{i,\,j\mathcal{J}} = \\ &= (\widetilde{\rho}_{ia\mathcal{A}}|\rho_{j\mathcal{J}b\mathcal{B}}) + (\rho_{ia\mathcal{A}}|\widetilde{\rho}_{j\mathcal{J}b\mathcal{B}}) - (\widetilde{\rho}_{ia\mathcal{A}}|\widetilde{\rho}_{j\mathcal{J}b\mathcal{B}}) = \\ &= \sum_{P\mathcal{P}} d_{ia\mathcal{A}}^{P\mathcal{P}} J_{j\mathcal{J}b\mathcal{B}}^{P\mathcal{P}} \end{aligned} \tag{11.44}$$

This expression is "robust" in the sense that it guarantees that the error in the approximated integrals is second order with respect to errors in the fitted densities. The last equality in (11.44) holds true provided that the same fitting domain is chosen for fitting both the $\rho_{ia\mathcal{A}}$ and $\rho_{j\mathcal{J}b\mathcal{B}}$ densities; in periodic systems this condition is fulfilled only if we formally consider an infinitely spread fitting basis set.

The auxiliary set (FF set) must be flexible enough to correctly reproduce PDs which hold angular momenta of higher order with respect to the AOs of the basis set itself. Pre-optimized large sets of GTFs are commonly used for this purpose and work is currently going on for obtaining well-performing FF sets. Morover, a scheme has been proposed by Manby, where the FF set, apart from few GTFs, basically consists of so-called Poisson Type Functions (PTF). For a given GTF orbital (labeled as G), the corresponding PTF, labeled as P, is defined as:

$$P(\mathbf{r}) = -\nabla^2/4\pi \; \cdot G(\mathbf{r}) \; . \tag{11.45}$$

PTFs hold neither charge nor multipoles of any order; furthermore, since Coulomb integrals involving PTFs transform to kinetic integrals, they are faster to compute and, what is especially important, vanish exponentially with distance. When used to fit charge densities which hold non-zero multipoles, PTFs must be complemented with GTFs of different angular momentum in order

to reproduce the multipoles; since PDs occurring in the LMP2 method have zero charge, s-type GTFs are not required.

In local-correlation methods, where the fitted densities are localized, also a local density fitting (LDF) technique is often adopted, where instead of the full set of the fitting functions only density-specific spatially restricted domains of the fitting functions (the so called fit-domains) are utilized. This technique is mandatory to achieve linear asymptotic scaling of the computational cost with the system size. This corresponds to the direct space fitting, activated by the keyword **DIRECT** in the **DFITTING** section.

For each set of PDs, having the same i index and $a\mathcal{A}$ on the same atom, a specific fit domain is defined. For accurate fitting the fit-domains should be chosen such that the corresponding fitting functions would provide a sufficient support to the fitted densities. Therefore we need a simple and at the same time stable technique for estimating the population of the density to be fitted on an atom. We define a quasi-population $q_{D\mathcal{D}}^{ia\mathcal{A}}$ of the product density $\rho_{ia\mathcal{A}}$ on an atom D in cell $\mathcal{D}$ as:

$$q_{D\mathcal{D}}^{ia\mathcal{A}} = \sum_{\mu\in D\mathcal{D}} \left(\sum_{\nu\in D\mathcal{D}} C_{\mu\, i} S_{\mu\nu} C_{\nu\, a\mathcal{A}} \right)^2 , \tag{11.46}$$

where $C_{\mu\, a\mathcal{A}}$ and $C_{\nu i}$ are the LCAO coefficients of PAOs and WFs, respectively, and $\mathbf{S}$ is the AO overlap matrix. Such a definition of the population slightly differs from the usual Mulliken form. The latter is not appropriate for our needs because the charge of any product density ρ_{ia} is zero due to the strict orthogonality between occupied and virtual spaces. A fit-domain in our approach comprises a certain number $N_{D\mathcal{D}}$ of atoms $D\mathcal{D}$ with the highest populations $q_{D\mathcal{D}}^{ia\mathcal{A}}$ for a block of PAOs belonging to a common center. As will be shown below, setting $N_{D\mathcal{D}}$ to 10-15 atoms is usually sufficient, but can be tuned through the **NMINCENT** keyword. Fitting domains so defined are finally adapted to fulfill symmetry invariance properties of the corresponding PD.

The approach activated by the **KSPACE** keyword is different, and is based on the reformulation of equation (11.44) in reciprocal space. As a first step we need to define a mesh of $\mathbf{k}$ points to sample the reciprocal space and obtain the Fourier Transform (FT) of the quantities of interest. This means dropping the lattice vector index $\mathcal{P}$ and dealing for each $\mathbf{k}$ with matrices having a size equal to the number of FFs in the reference cell:

$$\begin{aligned} \widetilde{K}_{a\mathcal{A},\, b\mathcal{B}}^{i,\, j\mathcal{J}} &= \sum_{\mathbf{k}} \widetilde{K}_{a\mathcal{A},\, b\mathcal{B}\ominus\mathcal{J}}^{i,\, j}(\mathbf{k}) \exp(-\imath\mathbf{k}\cdot\mathbf{R}_J) = \\ &= \sum_{\mathbf{k}} \sum_{P} d_{ia\mathcal{A}}^{P}(\mathbf{k}) J_{jb\mathcal{B}\ominus\mathcal{J}}^{P}(\mathbf{k})\ \exp(-\imath\mathbf{k}\cdot\mathbf{R}_J) \end{aligned} \tag{11.47}$$

where $d_{ia\mathcal{A}}^{P}(\mathbf{k})$ are the fitting coefficients in the reciprocal space:

$$d_{ia\mathcal{A}}^{P}(\mathbf{k}) = \sum_{Q} [J(\mathbf{k})]_{PQ}^{-1} J_{ia\mathcal{A}}^{Q}(\mathbf{k}), \tag{11.48}$$

while $J_{ia\mathcal{A}}^{P}(\mathbf{k})$ and $J_{PQ}(\mathbf{k})$ are the Fourier-images of the integrals (11.43). Thus, the fitting is performed in reciprocal space independently for each $\mathbf{k}$ point, and the approximated integrals are

finally obtained through a Back Fourier Transformation (BFT). The range of the summations for the FT step is defined for each WF according to a sphere centered on its centroid. The radius is set by default to 10 Å, and can be tuned through the **FITRAD** keyword.
A well known issue of the Coulomb series is their divergent behavior, which in the context of DFP emerges when evaluating the FT of integrals involving FFs of $p-$GTF type. Although the main fraction of the fitting basis set consists of Poisson functions, making the corresponding integrals exponentially decay and the Coulomb series thus quickly converge, one still needs a small number of p-type GTFs to represent the dipole moments of the fitted densities. The convergence of the series involving d- f- and g- GTFs is not divergent, but also rather slow, and thus are more conveniently treated in direct space as well. The same holds for the most diffuse part of the PTO basis (with exponent less than 1.0 Bohr^{-2}) which might cause numerical instabilities in the FT, and are also treated in the direct space together with the GTOs (this can be regulated via the **MIXEXP** keyword).

The basis set is divided into two parts: a rich PTF basis spanning the whole crystal, labeled as **P**, and a density-specific small basis, to fit the multipole moments, labeled as **G**. The PTF-fit is then performed in reciprocal space according to equations (**??**)-(**??**), while the GTF-based local fit is performed in direct space. However, a straightforward application of this scheme is not possible, since the metric matrix cannot be inverted either in direct or in reciprocal space (in the former case, one would need to transform the slowly decaying GTF-based integrals into the reciprocal space, while for the latter one would deal with a metric matrix of infinite dimensions due to the extended PTF-part of the basis). In order to decouple the two sets we project the GTF part of the auxiliary basis onto the space complementary to that spanned by the PTFs, using the Coulomb-projection operator:

$$|\Gamma) = |G) - \sum_{PP'} |P)[J^{\{P\}}]^{-1}_{P,P'}(P'|G) \tag{11.49}$$

The projected GTFs $\mathbf{\Gamma}$ can now be used as the direct space part of the fitting basis instead of the GTFs **G**. The block diagonal form of the metric matrix **J**, with zero off-diagonal blocks ($\mathbf{\Gamma}|\mathbf{P}$):

$$\mathbf{J} = \begin{array}{|c|c|}\hline (\mathbf{G}|\mathbf{G}) & (\mathbf{P}|\mathbf{G}) \\ \hline (\mathbf{P}|\mathbf{G}) & (\mathbf{P}|\mathbf{P}) \\ \hline \end{array} \Rightarrow \begin{array}{|c|c|}\hline (\mathbf{\Gamma}|\mathbf{\Gamma}) & \mathbf{0} \\ \hline \mathbf{0} & (\mathbf{P}|\mathbf{P}) \\ \hline \end{array}$$

now permits an independent inversion of both diagonal blocks: the PTF-block in the reciprocal space, and $\mathbf{\Gamma}$-block in the direct space. We have named this technique the direct-reciprocal-decoupled (DRD) DF.

The functions $\mathbf{\Gamma}$ serve as moment-fitting functions completely in the same manner as GTFs, since they differ from the latter only by a linear combination of momentless PTFs. Moreover, the quality of the fit does not deteriorate when shifting from the GTOs to the functions $\mathbf{\Gamma}$, since the projection (11.49) does not reduce or modify the space spanned by the initial fitting basis. The Γ-fit is performed in the same way as the direct-space fit of equation 11.44, but with the projected fitting

functions Γ instead of usual fitting functions and with much smaller fit-domains. Fit-domains are defined as in 11.46, but large domains are in fact not needed, since the fitting basis has a large reciprocal part which is not restricted to any local domain. For this reason, the corresponding default is set to 6 atoms in the case of a reciprocal space fit, which can be tuned as well using the **NMINCENT** keyword. We point out, that the opposite projection, *i.e.* the projection of the PTF part from the space spanned by GTFs (say $|\Pi) = |P) - \sum_{GG'} |G)[J^{\{G\}}]^{-1}_{G,G'}(G'|P))$ does not preserve translational invariance, which inhibits the reciprocal treatment of such functions.

Once the fit-domains are defined, one has to evaluate the integrals $(\mathbf{\Gamma}|\mathbf{\Gamma})$ and $(ia|\mathbf{\Gamma})$. Using the definition of the functions $\mathbf{\Gamma}$ (equation 11.49) yields:

$$(\Gamma|\Gamma) = (G|G) - \sum_{PP'} (G|P)[J^{\{P\}}]^{-1}_{P,P'}(P'|G) \tag{11.50}$$

and

$$(ia|\Gamma) = (ia|G) - \sum_{PP'} (ia|P)[J^{\{P\}}]^{-1}_{P,P'}(P'|G) \tag{11.51}$$

The first terms in the RHS of the expressions (11.50) and (11.51) are the usual integrals with the direct-space set of functions, while the second terms can be interpreted as the equivalent integrals, but fitted by the PTF-only fitting basis. As discussed above, this fit can be carried out only in the reciprocal space. Therefore, one has to employ the very same machinery of the PTF reciprocal-space fit as for the actual 4-index ERIs using equations (11.47) and (11.48). Summarizing :
(i) the set $\mathbf{\Gamma}$ it is a local set, since it is formed from the local $\mathbf{G}$ set and the same set but fitted by PTFs (see below).
(ii) the set $\mathbf{\Gamma}$ can carry multipole moments, since it differs from $\mathbf{G}$ only by some linear combinations of momentless PTFs, and
(iii) the robust fit with the set $\mathbf{\Gamma}$ can be performed independently from the PTF-fit, since the inversion of the $\mathbf{J}$ matrix is now conveniently factorized.
The latter allows us to subdivide the fitting of the integrals in two separate contributions:

$$\widetilde{K} = \widetilde{K}^P + \widetilde{K}^\Gamma \ . \tag{11.52}$$

The first term is the fit of the integrals through PTFs only, which constitute the major fraction of the auxiliary basis set and can be efficiently handled in reciprocal space without any convergence problems. The second term is the local fit and is processed in direct space.

Multipolar expansion for distant pairs

Consider a WF-PAO PD, $\rho_{ia\mathcal{A}}$, where the WF (i) is in the reference cell at site $\boldsymbol{r}_i$, and the PAO $(a\mathcal{A})$ belongs to the domain of i. The set $\{Q(i;a\mathcal{A})_{\ell,m}\}$ of all multipole moments of $\rho_{ia\mathcal{A}}$ with respect to $\boldsymbol{r}_i$ up to a maximum order $\overline{\ell}$, (set by default to 4 and modifiable via the **MULTIPO** keyword), is calculated once and for all; the zero-order moment is always zero, due to WF-PAO orthogonality. When two PDs $\rho_{ia\mathcal{A}}$, $\rho_{j\mathcal{J}b\mathcal{B}}$ are so far apart that they are essentially confined to separate spheres, the corresponding ERI can be approximated as a sum of interactions between their multipoles centered in $\boldsymbol{r}_i$ and $\boldsymbol{r}_j + \boldsymbol{R}_\mathcal{J}$, respectively:

$$K^{i,j\mathcal{J}}_{a\mathcal{A},b\mathcal{B}} \approx \sum_{\ell,m;\ell',m'} Q(i;a\mathcal{A})^{\ell,m}\, U^{i,j\mathcal{J}}_{\ell,m;\ell',m'} Q(j;b\mathcal{B}\ominus\mathcal{J})^{\ell',m'} \tag{11.53}$$

where $\mathbf{U}^{i,j\mathcal{J}}$ is an interaction matrix which depends only on the relative position $\boldsymbol{r}_j + \boldsymbol{R}_{\mathcal{J}} - \boldsymbol{r}_i$ of the two centers. The translational equivalence of PDs centered in different cells has here been exploited.

Note that the systematic use of the multipolar scheme for WF pairs beyond a distance d_{MP} requires that "cross excitations" from one WF to the domain of the other are neglected, because they always involve ERIs between overlapping PDs which cannot be described according to equation (11.53). This further approximation is justified by the fact that the norm of "cross PDs" exponentially decays with the distance between the two WFs.
The multipolar approximation allows a fast evaluation of integrals for distant pairs: the computational cost is insignificant and high accuracy is warranted. Our experience suggests that a suitable value for d_{MP} is 8 Å.

11.3 LMP2 correction to the Hartree-Fock Density Matrix

In a second-quantization formalism, the one-electron position-spin Density Matrix (DM) $\gamma(\mathbf{x};\mathbf{x}')$ associated with a normalized many-electron wavefunction, corresponding to the selected state $|\Psi\rangle$ can be written as follows:

$$\gamma(\mathbf{x};\mathbf{x}') \;=\; \sum_{PQ} \langle\Psi|a_P^\dagger\, a_Q|\Psi\rangle\; \Phi_P^\star(\mathbf{x}')\, \Phi_Q(\mathbf{x}) \tag{11.54}$$

Here reference is made to an orthonormal set of one-electron spin-orbitals, $\Phi_Q(\mathbf{x}) \equiv \Phi_Q(\mathbf{r},\omega)$, which span the Fock space in which $|\Psi\rangle$ is defined, and $a_Q^\dagger$, a_Q are the corresponding creation and annihilation operators. By integrating $\gamma(\mathbf{r},\omega;\mathbf{r}',\omega)$ over the spin coordinate ω, we get $R(\mathbf{r};\mathbf{r}')$, the so-called *position* DM. The *momentum* DM $P(\mathbf{p};\mathbf{p}')$ is just the six-dimensional Fourier transform of $R(\mathbf{r};\mathbf{r}')$.

With reference to the AO basis set $\{\chi_{\mu\mathbf{g}}(\mathbf{r})\}$ adopted in the Crystal and Cryscor codes, the position DM can always be written as follows:

$$R(\mathbf{r};\mathbf{r}') = \sum_{\mathbf{g},\mathbf{l}} \sum_{\mu\nu} P_{\mu\nu\mathbf{g}}\; \chi_{\mu\mathbf{l}}(\mathbf{r})\; \chi_{\nu(\mathbf{l}+\mathbf{g})}(\mathbf{r}') \tag{11.55}$$

where the translational invariance of the lattice has been used.

In the next two subsections we first recall how this matrix is obtained from a single-determinantal solution $|\,\Psi^{\mathrm{HF}}\rangle$ or $|\,\Psi^{\mathrm{DFT}}\rangle$, then we show how to calculate the MP2 correction to its HF estimate. The discussion is restricted to closed-shell, non-conducting crystals.

Density Matrix from single-determinantal wavefunctions

The single-determinantal "X" wavefunction ($X = \mathrm{HF}$, KS-DFT) can be expressed in two equivalent ways:

$$\begin{aligned} |\Psi^X\rangle \;\;&\longleftrightarrow\;\; ||\cdots\, \psi^X_{j,\kappa}\alpha\;\; \psi^X_{j,\kappa'}\beta\;\cdots|| \;= \\ &=\;\; ||\cdots\, w^X_{i,\mathbf{g}}\alpha\;\; w^X_{i,\mathbf{g}}\beta\;\cdots|| \end{aligned} \tag{11.56}$$

In the first line, the Slater determinant is constructed with the canonical doubly-occupied crystalline orbitals (CO) that satisfy the self-consistent-field equations:

$$\widehat{h}^X \; \psi^X_{j,\kappa}(\mathbf{r}) \;=\; \varepsilon^X_{j,\kappa} \; \psi^X_{j,\kappa}(\mathbf{r}) \tag{11.57}$$

$$\psi^X_{j,\kappa}(\mathbf{r}) \;=\; \sum_{\mu} a^X_{j,\kappa;\mu} \left[\sum_{\mathbf{g}} e^{\imath\,\kappa\cdot\mathbf{g}} \; \chi_{\mu\mathbf{g}}(\mathbf{r}) \right] \tag{11.58}$$

The band index j runs from 1 to $N_0/2$, while the wavevector κ is one of the L vectors in the first Brillouin Zone of reciprocal space which form the Monkhorst net associated with the selected cyclic crystal. In the second line, the same determinant is expressed using the Wannier functions, labelled by $\mathbf{g}$, one of the L direct lattice vectors of the cyclic crystal, and by the in-cell index i running from 1 to $N_0/2$; the WFs are real-valued, well localized, symmetry adapted functions of $\mathbf{r}$, [5, 6] which span altogether the same space as the occupied COs and are translationally equivalent and mutually orthonormal:

$$w^X_{i,\mathbf{0}}(\mathbf{r}) = \; w^X_{i,\mathbf{g}}(\mathbf{r}+\mathbf{g}) \qquad ; \tag{11.59}$$

$$\int \; w^X_{i,\mathbf{g}}(\mathbf{r}) \, w^X_{i',\mathbf{g}'}(\mathbf{r}) \; d\mathbf{r} = \delta_{ii'}\delta_{\mathbf{g}\mathbf{g}'} \tag{11.60}$$

Using equation (11.54) with $|\Psi\rangle = |\Psi^X\rangle$ and making reference to the orthonormal sets $\{\Phi_Q(\mathbf{r},\omega)\}$ of the COs or of the WFs, gives us, after integration over spin:

$$R^X(\mathbf{r};\mathbf{r}') \;=\; 2 \sum_{j=1}^{N_0/2} \sum_{\kappa} \psi^X_{j,\kappa}(\mathbf{r}) \; [\psi^X_{j,\kappa}(\mathbf{r}')]^* \tag{11.61}$$

$$\;=\; 2 \sum_{i=1}^{N_0/2} \sum_{\mathbf{g}} w^X_{i,\mathbf{g}}(\mathbf{r}) \; w^X_{i,\mathbf{g}}(\mathbf{r}') \tag{11.62}$$

Substitution of equation (11.58) into equation (11.61) and comparison with equation (11.55) gives immediately:

$$P^X_{\mu\nu\mathbf{g}} \;=\; 2 \sum_{j=1}^{N_0/2} \sum_{\kappa} e^{-\imath\,\kappa\cdot\mathbf{g}} \left[a^X_{j,\kappa;\mu} \; \left(a^X_{j,\kappa;\nu}\right)^* \right] \tag{11.63}$$

The MP2 correction to the HF Density Matrix

The MP2 level of theory provides a first-order approximation of the correlated ground-state wavefunction:

$$|\Psi'\rangle \;=\; \zeta \; \left(|\, \Psi^{\mathrm{HF}}\rangle + |\Psi^{(1)}\rangle \right) \quad , \tag{11.64}$$

where ζ is the appropriate normalization factor. A technique has been implemented in Cryscor which is based on a "Lagrangian" approach, according to which the DM provides the first order response of the system energy E to an arbitrary external one-electron perturbation $\alpha\,\widehat{X}$, with α being the strength parameter, and the one-electron operator $\widehat{X}$ designated in the first-quantization

as $\hat{x}^c$. If $|\Psi\rangle$ were the outcome of a *variational* calculation performed for the unperturbed system, the Hellmann-Feynmann theorem:

$$\begin{aligned} \left.\frac{d\,E}{d\,\alpha}\right|_{\alpha=0} &= \langle\Psi|\widehat{X}|\Psi\rangle = \\ &= \sum_{PQ} \langle\Psi|a_P^\dagger\, a_Q|\Psi\rangle \int \Phi_P^\star(\mathbf{x})\,\hat{x}^c\Phi_Q(\mathbf{x})\,d\mathbf{x} \end{aligned} \tag{11.65}$$

would hold. Comparison of (11.65) with equation (11.54) allows us to identify here the elements R_{PQ} of the density matrix in the $\{\Phi_Q\}$ representation as the weighting factors for the matrix elements of the perturbation operator $X_{PQ} = \int \Phi_P^\star(\mathbf{x})\,\hat{x}^c\,\Phi_Q(\mathbf{x})\,d\mathbf{x}$ in the energy derivative expression. In a *non-variational* treatment, as is the case for MPn or standard CC methods, the energy still can be written via a variational expression, by setting up the corresponding Lagrangian. However, in this case, in addition to the excitation amplitudes which are variational parameters also in the standard variational approach, a second set of parameters is to be varied. These parameters are the Lagrange multipliers, corresponding to the MPn or CC equations used as the constraints in the Lagrangian.

In the case of MP2, the Lagrangian is nothing else than the well-known Hylleraas functional. Only the doubles amplitudes t_{ab}^{ij}, corresponding to the excitations from a pair of occupied orbitals (i,j) to a pair of virtual ones (a,b), enter the MP2 formalism. Due to the symmetric form of the Hylleraas functional, the Lagrange multipliers $\overline{t}_{ab}^{ij}$ turn out to be not independent variational parameters, but rather the contravariant amplitudes: $\overline{t}_{ab}^{ij} = 2t_{ab}^{ij} - t_{ba}^{ij}$.

Next, from the resulting expression of $dE/d\alpha|_{\alpha=0}$, the weighting factors for the X_{PQ} matrix elements in analogy to (11.65) can be defined as the elements of the DM R_{PQ}. In case of real orthonormal orbitals $\{\phi_i(\mathbf{r})\}$, $\{\xi_a(\mathbf{r})\}$ which span the occupied and virtual HF manifold, respectively, the following expression for the MP2 correction to the HF position DM is obtained:

$$\begin{aligned} R^{\mathrm{MP2}}(\mathbf{r};\mathbf{r}') &= -2\sum_{kab} \overline{t}_{ab}^{jk}\, t_{ab}^{ik}\,\phi_i(\mathbf{r})\,\phi_j(\mathbf{r}') \;+ \\ &\quad +2\sum_{cij} \overline{t}_{bc}^{ij}\, t_{ac}^{ij}\,\xi_a(\mathbf{r})\,\xi_b(\mathbf{r}')\;, \end{aligned} \tag{11.66}$$

This formula can be generalized to the periodic case with the occupied space spanned by Wannier functions, and the virtual space truncated according to the local approximation and represented by nonorthogonal projected atomic orbitals.

An advantage of this approach is that from the size-extensive MP2 expression for the correlation energy a size-extensive correlation correction to the HF DM is obtained. Besides, its implementation turns out to be computationally more efficient than that of $\langle\Psi''|a_P^\dagger\, a_Q|\Psi''\rangle$. It must be pointed out that the Lagrangian DM formalism allows for a further improvement of the DM, in particular by including in the Lagrangian the constraints required for the validity of the Brillouin theorem in the presence of a perturbation, *i.e* the "orbital relaxation". Presently the approach, implemented in CRYSCOR is orbital-unrelaxed, but the work on inclusion of the orbital relaxation effects in the DM is underway.

Bibliography

[1] Pisani, C., Busso, M., Capecchi, G., Casassa, S., Dovesi, R., Maschio, L., Zicovich-Wilson, C., and Schütz, M. *J. Chem. Phys.* **122**, 094133 (2005).

[2] Pisani, C., Maschio, L., Casassa, S., Halo, M., Schütz, M., and Usvyat, D. *J. Comp. Chem.* **29**, 2113 (2008).

[3] Dovesi, R., Saunders, V. R., Roetti, C., Orlando, R., Zicovich-Wilson, C. M., Pascale, F., Doll, K., Harrison, N. M., Civalleri, B., Bush, I. J., D'Arco, P., and Llunell, M. *CRYSTAL09 User's Manual*. Università di Torino, Torino, (2009). http://www.crystal.unito.it.

[4] Wannier, G. H. *Phys. Rev.* **52**, 191 (1937).

[5] Zicovich-Wilson, C. M., Dovesi, R., and Saunders, V. R. *J. Chem. Phys.* **115**, 9708 (2001).

[6] Casassa, S., Zicovich-Wilson, C. M., and Pisani, C. *Theor. Chem. Acc.* **116**, 726 (2006).

[7] Pulay, P. *Chem. Phys. Lett.* **100**, 151 (1983).

[8] Pulay, P. and Saebø, S. *Theor. Chim. Acta* **69**, 357 (1986).

[9] Werner, H.-J., Knowles, P. J., Lindh, R., Manby, F. R., Schütz, M., Celani, P., Korona, T., Rauhut, G., Amos, R. D., Bernhardsson, A., Berning, A., Cooper, D. L., Deegan, M. J. O., Dobbyn, A. J., Eckert, F., Hampel, C., Hetzer, G., Lloyd, A. W., McNicholas, S. J., Meyer, W., Mura, M. E., Nicklass, A., Palmieri, P., Pitzer, R., Schumann, U., Stoll, H., Stone, A. J., Tarroni, R., and Thorsteinsson, T. *MOLPRO: a package of ab initio programs, version 2006.1*, (2006). http://www.molpro.net.

[10] Usvyat, D. *Adsorption of molecules on surfaces with* CRYSCOR09 , http://www.cryscor.unito.it (2010).

[11] Maschio, L., Usvyat, D., Manby, F., Casassa, S., Pisani, C., and Schütz, M. *Phys. Rev. B* **76**, 075101 (2007).

[12] Usvyat, D., Maschio, L., Manby, F., Casassa, S., Schütz, M., and Pisani, C. *Phys. Rev. B* **76**, 075102 (2007).

[13] Maschio, L. and Usvyat, D. *Phys. Rev. B* **78**, 073102 (2008).

[14] Usvyat, D. and Schütz, M. *J. of Phys.: Conf. Series* **117**, 012027 (2008).

[15] Saebø, S. and Pulay, P. *J. Chem. Phys.* **86**, 914 (1987).

[16] Schütz, M., Lindh, R., and Werner, H.-J. *Mol. Phys.* **96**, 719 (1999).

[17] Schütz, M., Hetzer, G., and Werner, H.-J. *J. Chem. Phys.* **111**, 5691 (1999).

[18] Boys, S. F. In *Quantum Theory of Atoms, Molecules, and the Solid State,* Löwdin, P. O., editor, 253. Academic, New York (1996).

[19] Pipek, J. and Mezey, P. G. *J. Chem. Phys.* **90**, 4916 (1989).

[20] Mulliken, R. S. *J. Chem. Phys.* **23**, 1833 (1955).

[21] Mulliken, R. S. *J. Chem. Phys.* **23**, 1841 (1955).

[22] Boughton, J. W. and Pulay, P. *J. Comp. Chem.* **14-6**, 736 (1993).

Index

ALLEQ, 23
AMPREA, 30
ASYMDOM, 25
ATOMS, 35

Bugs, 8

Citing Cryscor, 7
CONTOL, 30, 42, 47
Conventions, 5
Convergence criteria, 30
COORDINA, 35
COREQ, 23

DELPAO, 23
Density Fitting, 26
Density matrix, 32
DENSMAT, 32, 40, 49
DFITTING, 14, 26
DFSCHW, 28, 29
DIRECT, 26
Domains, 19
DOMCOE, 20, 22
DOMDEF, 14, 20, 21, 40, 44, 64
DOMDEF2, 20, 21
DOMMOL, 14, 20, 21, 64
DOMONLY, 24
DOMPUL, 14, 20, 64
DOMREA, 24
DSCREEN, 25
DTIME, 36
Dual basis set, 31
DUALBAS, 31, 48

END, 14, 16
ENDDF, 27
ENVPAIR, 19

Error sources, 37
Excitation domains, 19
External units, 7

FITRAD, 29
FIXDOM, 24
FIXINDEX, 24
Fixindex procedure, 23
FIXNPAOS, 24
FIXPAIR, 24
FIXWF, 24
Freezing the integrals, 23
Functionality, 6

General keywords, 16
Geometry optimization, 23
GUESDUAL, 31

How to run, 7

Input format, 14
Integrals, 25
INTREA, 30

KNET, 14, 16
KSPACE, 26

LENJONES, 18, 48, 64, 71
Lennard-Jones extrapolation, 18
Local periodic MP2 theory, 57
LOCALWF, 9
LONTOL, 23, 61

MAXDOM, 20
MEMORY, 14, 16
METHDIR, 29
METHREC, 29
MINPOP, 17, 18, 63

MIXEXP, 27
MOENPAIR, 19
MOLATOMS, 19
MOLPAIR, 19
Monkhorst net, 16
MP2 equations, 30
MPREAD, 26
MULTIPO, 26, 75
Multipolar Expansion, 25

NCONSTR, 29
NMINCENT, 28, 29
NONEQ, 23
NPAOSREA, 24

Occupied space description, 17
Output, 39

PAIR, 18, 42, 45
Pair partitioning, 19
Pair selection, 18
PAIRREA, 24
Plotting options, 34
PMP2, 32, 50
PRINPLOT, 35, 39
PRINTBAS, 27, 28
Printing options, 34

RDFMWF, 7
Restart, 30

SCF cycle, 30
SCHWARZ, 25, 28
Script, 7
SYMMWF, 9

TESTDOM, 22
The localization step, 9
THRSHEIG, 29
Timing, 36
TOBJ, 23, 42, 59
TOLEAMPS, 30
TOLEQPRS, 30
TOLMP2, 37
Truncation criteria, 23
Two-electron integrals, 25

USRBASIS, 27

VALENCE, 9
Virtual space description, 19