

# GNXAS: A SOFTWARE PACKAGE FOR ADVANCED EXAFS MULTIPLE-SCATTERING CALCULATIONS AND DATA-ANALYSIS

ADRIANO FILIPPONI<sup>1</sup> AND ANDREA DI CICCIO<sup>2</sup>

<sup>1</sup>*INFN, Dipartimento di Fisica,  
Università degli studi dell'Aquila, via Vetoio,  
67100 Coppito (Aq), Italy  
adriano.filipponi@aquila.infn.it*

<sup>2</sup>*INFN, Dipartimento di Matematica e Fisica,  
Università di Camerino, via Madonna delle Carceri,  
62032 Camerino (MC), Italy  
dicicco@unixas.unicam.it*

(Received 20 August 2000)

**Abstract:** Computational and theoretical advances for the interpretation of the x-ray absorption spectroscopy (XAS) cross-section, and associated extended x-ray absorption file structure (EXAFS), originally published in Phys. Rev. B **52**, 15122 and 15135, (1995), are at the basis of an advanced software package for EXAFS data analysis, denominated GNXAS. This software package is composed of a series FORTRAN programs for EXAFS data-analysis (including the main codes `crymol`, `phagen`, `gnpeak`, `gnxas`, and `fitheo`) together with several utility programs for sample optimization, automatic background subtraction, and edge analysis. The GNXAS package allows scientists to perform accurate multiple-scattering EXAFS calculations of the XAS cross-section achieving a reliable refinement of the structural models in the short-range. These characteristics are useful to take full advantage of the high data quality obtained at modern synchrotron radiation facilities. The philosophy, methodology, and practical usage aspects of the main data-analysis codes are described in details in this paper. The procedures for the estimation of statistical errors and for the data analysis of liquid and highly disordered systems are also discussed. A useful list of references on the development of the method and applications to various scientific fields is included. The content of this paper can be regarded as a full documentation and guide for proper usage of the current version of the GNXAS software.

**Keywords:** EXAFS, XAS, multiple-scattering, GNXAS, structure

## 1. Introduction

The x-ray absorption spectroscopy (XAS) is a powerful structural technique for investigating the short-range environment around selected atomic species in condensed matter. This insight can be obtained from the analysis of the quantum interference pattern, usually detectable above any inner shell absorption edge, generated by the scattering of the photoelectron on the potential of the surrounding atoms. These oscillations are usually

referred to as x-ray absorption fine structure (XAFS) or EXAFS. X-ray absorption spectra can be collected in transmission geometry or using other secondary emission effects, (fluorescence, electron yield...). The collection of a spectrum as a function of the x-ray photon energy  $E$  allows one to measure, above the continuum threshold  $E_e$ , the EXAFS  $\chi(k)$  as a function of the photoelectron momentum  $k = \sqrt{2m(E - E_e)}/\hbar$ . The applications of this spectroscopy cover a wide range of scientific disciplines ranging from physics, chemistry, biology and earth sciences. XAS experiments are possible using laboratory sources but the most efficient and accurate experiments are performed, since the 80's, using synchrotron radiation. At present several tens of beamlines exist all around the world where it is possible to perform high quality XAS (EXAFS) experiments in various energy domains. These international facilities are available for usage to a wide scientific community.

In order to facilitate the scientists in the preparation of their experiments and in the subsequent data analysis phase, we have initiated since the year 1990, an ambitious software development project. The focus of this project was on the possibility to apply EXAFS data analysis to disordered systems and consequently to link the XAS signal to features of the  $n$ -atom distribution functions  $g_n$ . This is the origin of the name GNXAS. The GNXAS suite of programs was developed around an advanced software for EXAFS data analysis based on multiple-scattering (MS) calculations and a rigorous fitting procedure of the raw experimental data [1–3].

GNXAS is the result of over 10 years of research efforts in the field of x-ray absorption spectroscopy and several scientists have contributed to its development. Details concerning the historical developments can be found in Appendix A1 where proper credit is given to people contributing in various ways to the project. This software has been widely tested and has been applied to a variety of systems including simple gas-phase molecules, clusters, nanocrystals, molecular solids and liquids, crystalline and liquid metals, solid and molten salts, amorphous solids and glasses, solutions, high- $T_c$  superconductors, and biological matter (see current GNXAS publication list in Appendix A2).

The present software distribution (see Appendix A3) includes the main data-analysis programs `crymol`, `phagen`, `gnpeak`, `gnxas`, and `fittheo` as well as several utility programs for sample optimization and characterization, automatic background subtraction, and edge fitting/deconvolution.

The main characteristics of the GNXAS data-analysis software are:

- 1) atomic phase shifts calculations in the muffin-tin approximation are based on atomic self-consistent relativistic calculations. Account for the neighbors is taken.
- 2) Inclusion of inelastic losses through complex Hedin-Lundqvist potential.
- 3) Calculation of MS signals associated with two, three, and four atom configurations using advanced algorithms.
- 4) Use of an advanced fitting procedure that allows one;
  - a) to fit simultaneously several spectra containing several edges,
  - b) to use directly the raw data without any pre-analysis,
  - c) to account for complex background multi-electron excitation features,
  - d) to use various model peaks for the pair, triplet and quadruplet distribution functions, including non Gaussian models and extremal cases. In all cases absolute values of the structural parameters can be refined.

- e) To treat liquid phase or disordered systems and extract reliable  $g(r)$  functions in the short-range limit ( $\leq 4 \text{ \AA}$ ).
  - f) To perform a rigorous statistical error analysis and plot two-dimensional correlation maps.
- 5) To provide a flexible scientific tool for EXAFS data analysis where the user has access to every stage of the calculation. In this sense the GNXAS package is not a black box.
  - 6) Full modularity makes it easy to interface parts of the GNXAS software with other available software.

All these characteristics make the GNXAS package a very useful tool for performing a modern EXAFS data-analysis and to extract accurately all the valuable information contained in modern high-quality x-ray absorption data. The algorithms have been developed and coded directly by the main scientists involved in x-ray absorption spectroscopy research and are continuously improved and updated. This has the advantage of having access to a living research tool. On the other hand the programs are sometimes characterized by a poor user interface and mistakes in input parameters, leading to unphysical results, are often not recognized by the programs. This explicit choice of the authors requires users willing to acquire or having already sufficient knowledge of the underlying theoretical background of the EXAFS spectroscopy, in order to avoid trivial mistakes and program misuse. It should be said here that other packages for MS EXAFS data analysis, that may appear more user friendly at a first glance, hide similar traps since the complexity resides in the x-ray absorption spectroscopy calculation and not in the particular code being used.

The purpose of this extended documentation is to inform, as completely as possible, the scientists interested in the usage of the GNXAS package. Users are also encouraged to read and cite the extensive original literature and the numerous applications listed in the publication list contained in Appendix A2. The paper is organised as follows: an overview of the GNXAS package is given in Section 2; the detailed description of `crymol`, `phagen`, `gnpeak`, `gnxas`, and `fitheo` programs is given in the successive Sections 3, 4, 5, 6, and 7, respectively; discussion about the estimation of statistical errors and the data-analysis of highly disordered solid and liquid structures (`grfit`, `grxas`, and `grrec`) are given in sections 8 and 9, respectively.

## 2. Overview of the package

The GNXAS package includes the EXAFS data-analysis programs, which constitute the main part of the package and are among the most advanced codes available, and several other utility programs for sample optimization, automatic background subtraction, and edge analysis.

For the EXAFS data analysis software the backbone of the GNXAS package is composed of three main codes, in the logical sequence:

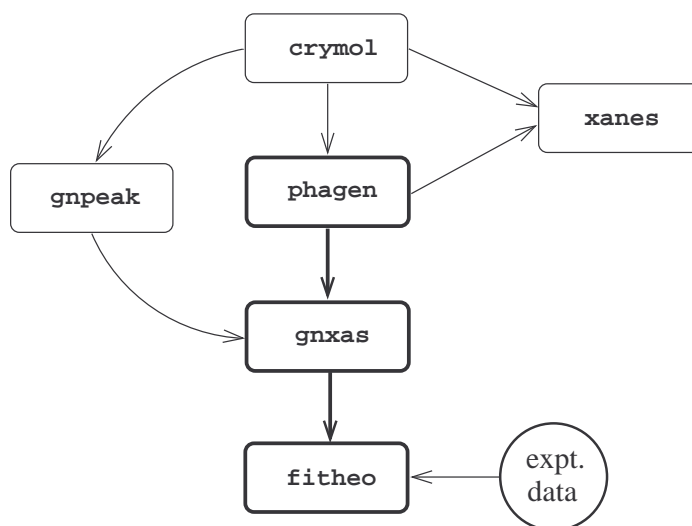
- 1) `phagen`, potential and phase shift generation (see Section 4).
- 2) `gnxas`, signal calculation for reference geometries (see Section 6).
- 3) `fitheo`, advanced fitting of the experimental data (see Section 7).

Two other programs are able to prepare automatically the input for `phagen` and `gnxas` starting directly from model molecular positions or crystallographic data. These are

particularly useful in the case of complex structures where the configuration counting is not trivial. These programs are:

- 4) `crymol`, allows one to treat complex molecular and crystallographic structures providing input information for `phagen`, `gnpeak`, and `xanes`; the program identifies prototypical phase-shift atoms and select a suitable cluster of atoms for successive XAS calculations (see Section 3).
- 5) `gnpeak`, is based on a general algorithm able to identify inequivalent two-body, three-body and four-body configurations in various structures providing valuable input information for `gnxas` (see Section 5).

A simplified logical flow-chart related to these data analysis programs is illustrated in the figure.



Proper XANES calculation should adopt a self consistent non-muffin tin potential, nevertheless users may be interested to test the performance of the phase shifts for muffin-tin potential as generated by `phagen` on the near-edge region. For this purpose an additional program is also supplied:

- 6) `xanes`, in which the continued fraction algorithm is fully implemented for non-symmetrized XANES calculation. For the reasons mentioned above the usage of this program for scientific applications is discouraged.

Several other programs were written to facilitate the output readout, in particular:

- 7) `phaplo`, produces a graphic output for the phase-shift file generated by `phagen` (see Section 4).
- 8) `reader`, converts the binary signal files generated by `gnxas` into ASCII files to facilitate visualizing and checking the output (see Section 6).
- 9) `contour`, to convert character contour plots in the `fitheo` output and facilitates error analysis (see Section 8).

Three additional programs have been written to facilitate the analysis of disordered systems (including liquids) where the structure is described in terms of radial distribution functions  $g(r)$  (see Section 9). These are:

- 10) `grfit`, fitting utility of model radial distribution function data.
- 11) `grxas`, calculates the EXAFS signal associated with a given partial radial distribution function.
- 12) `grrec`, allows reconstruction of  $g(r)$  after model refinement.

Finally the GNXAS distribution (see Appendix A3) includes miscellaneous programs useful for XAS based research:

- 13) `xasam`, a program for sample thickness optimization in the transmission geometry.
- 14) `jesf`, an automatic routine for EXAFS extraction, Fourier transformation, and noise evaluation. This program is useful for a qualitative data-analysis and it is routinely and successfully used at the ESRF-BM29 beamline for on-line data-analysis and EXAFS visualization during the acquisition of experimental data.
- 15) `deconv`, a deconvolution program suitable to deconvolve the core-hole life-time broadening in high quality spectra.

### 3. Program `crymol`

#### 3.1. Introduction

The program `crymol` is able to manipulate the geometrical information of a model structure in order to generate automatically the input files for the programs `phagen` and `gnpeak`. The program performs two main actions:

- a) It generates a cluster of sufficient size to count, with the correct multiplicity, all of the pair, triplet and four-atoms configurations associated with any non-equivalent photoabsorber up to a given cut-off distance (in this way one can account for all the multiple scattering paths involving up to four atoms).
- b) Select the different types of atoms that differ in atomic number, types of neighbors and distances within a given tolerance and build, for each case, appropriate mini-clusters to be used in the phase-shifts calculation.

The input geometrical information includes both molecular or crystalline structure and chemical bond information. Three structural input options are available:

- 1) Molecular case (option XYZ): atomic positions specified in Cartesian coordinates.
- 2) Simple crystalline structure (option CRY) specified by the unit cell shape and fractional positions of the atoms.
- 3) Complex crystalline structure (option SYM) where explicit use of the symmetries is made.

In the crystal cases an important input parameter is the size of the region that reproduces correctly the structure around a selected x-ray photoabsorbing atom. The central unit cell is surrounded symmetrically by an appropriate number of cells and the atoms in the final cluster are chosen in such a way that each photoabsorber in the central cell is surrounded by the correct geometry up to the given cutoff distance. This gives also a constraint to the frequency of the paths which will be counted correctly in the successive path analysis (see `gnpeak` documentation for details).

Chemically bonded neighbors are defined through a maximum distance table. This neighbor information is retained in the cluster generation. In the case of unconnected structures such as layered crystals, molecular crystals, *etc.* it is possible to chose whether to include or exclude unconnected atoms.

An output file containing the mini-cluster information for the phase-shift calculation (suitable input for the program `phagen`) is generated. For every inequivalent atom three different phase shifts calculations can be in principle required according to the following cases:

- (a) The atom is the photoabsorber (excited) one.
- (b) The atom is neutral and surrounded by neutral atoms.
- (c) The atom is neutral, but it is surrounded by one excited photoabsorber atom.

The program selects which of the above cases are appropriate for each type of atom in the structure, the difference among cases (b) and (c) can be made inactive.

### 3.2. *Input command file*

`crymo1` usually runs using an input command file containing the relevant structural information and input parameters. Detailed description of the `crymo1` input cards follows, indicating the proper format when the free format is not used.

#### 3.2.1. *Detailed card description*

##### 1. ICM (A3)

Selects between three possible types of geometrical input:

- a) ICM='XYZ' (if ICM='XYZ' see #.a choice)  
Atomic positions specified in Cartesian coordinates (useful for molecules).
- b) ICM='CRY' (if ICM='CRY' see #.b choice)  
Simple crystals: the unit cell structure and the fractional coordinates of all atoms are required.
- c) ICM='SYM' (if ICM='SYM' see #.b and #.c choices)  
Complex crystals: the input requires the atomic position of the inequivalent atoms and the set of symmetry operations for a generic site (defined according to the "International Tables For Crystallography"). Atoms located in sites with special symmetry are automatically recognized.

##### 2. NOME (character string, max A20).

Name identifying the crystalline structure.

NOTE: cases b and c (ICM='CRY' or ICM='SYM') crystallographic input only!

##### 2.1.b. XTAL (character\*1)

Crystallographic unit cell type:

T = Triclinic  
C = Cubic  
H = Hexagonal  
M = Monoclinic  
O = Orthorombic  
R = Trigonal  
E = Tetragonal

##### 2.2.b. a, b, c

Unit cell dimensions.

If T='C', 'R' type only a (a=b=c).

If T='E' or T='H' type only a, c (a=b).

2.3.b.  $\alpha, \beta, \gamma$ 

Angles of the unit cell. This card does not appear if T='C', 'E', 'O' (all angles are 90°) and in the case T='H' ( $\alpha = \beta = 90^\circ$ ,  $\gamma = 120^\circ$ ).

In the trigonal case T='R' only  $\alpha$  is needed ( $< 120^\circ$ ).

In the monoclinic case T='M' only  $\beta$  is needed ( $\alpha = \gamma = 90^\circ$ ).

## 3. NA

Represents the total number of (inequivalent) atoms in the molecule or unit cell.

## 4. I, SYM, X(I), Y(I), Z(I), NV(I), CHARGE(I), RS(I)

Specific information on each atom of the molecule or unit cell. The input file contains NA cards of this type. Parameters CHARGE(I), RS(I) do not affect the crymol run, they are just copied on the input file for phagen.

I Sequential number labeling each atom of the molecule or unit-cell (I=1, 2, ..., NA).

SYM The (standard) chemical symbol of atom I (e.g. 'Si', 'C').

X, Y, Z(I) Coordinates of the atom: for (ICM='XYZ') they are Cartesian coordinates in (Å) whereas in the crystalline input cases (ICM='CRY' or ICM='SYM') they are the dimensionless fractional coordinates in the unit cell.

NV(I) Known (or guessed) number of first neighbors.

CHARGE(I) Fractional electronic charge (put 0., not yet implemented in phagen).

RS(I) Muffin-tin radius of the atom. (WARNING: Value in atomic units! 1 a.u.=0.529 Å).

NOTE: The following cards 4.1, 4.2, 4.3, and 4.4 refer to the case c (ICM='SYM') only.

## 4.1.c NXYZA

Number of pure translation symmetry operations  $NXYZA \geq 1$ .

## 4.2.c XYZA(N, 1), XYZA(N, 2), XYZA(N, 3)

Components of the translation in fractional units of the elementary cell. At least one (trivial) translation by the null vector has to be specified as "0.,0.,0.". A translation of  $0.5\vec{a}$  and  $0.25\vec{b}$  is indicated as "0.5,0.25,0.". NXYZA cards of this type must appear.

## 4.3.c NOP

Number of distinct rotational-translational symmetry operation, for a generic point, according to the space group of the crystal.

## 4.4.c M, CO(M, 1), CO(M, 2), CO(M, 3), TT(M, 1), TT(M, 2), TT(M, 3)

Rotational and translational information on the symmetry operation. NOP cards of this type are required.

M sequential number labeling the operation (has no effect).

CO(M, #) are 3 (A6) format variables defining the rotation operation. The three fields refer to the x, y, and z coordinates, respectively. Each (A6) field is subdivided into three (A2) subfields. The even positioned characters are either blank if unused or equal to x (second), y (fourth), or z (sixth). At least one of them has to be not-blank. The first,

third, and fifth characters can be either blank (positive or unused) or – (negative).

TT(M, #) defines the additional possible translation to the symmetry operation.

For example, the symmetry operation  $(-y, x-y, z+1/3)$  is indicated as:

#, ' -y ', ' x-y ', ' z ', 0., 0., 0.33333

#### 5. NEI (character\*1), NEU (character\*1), FNEU (character\*1)

Algorithm options: recommended choice is “YNY”:

NEI='Y' the neighbor search within maximum distance limits specified by the cards (6.-7.) proceeds until NV(I) (cards 4.) neighbors are found. If other atoms happen to be within the first neighbor range they are just ignored. If the neighbors are less than NV(I) an error message appears. This option is useful to check the input, however, if the number of neighbors is unknown the alternative option may be useful.

NEI='N' all the atoms within maximum distance limits specified by the cards (6.-7.) are labeled as first neighbors. WARNING: there is a dimension limit of 12 for the number of first neighbors.

NEU='Y' only atoms connected by first neighbor bonds are included in the final cluster.

NEU='N' all atoms up to the cutoff-distance are included in the final cluster. In the case of molecular crystals or layered structures the NEU flag gives the flexibility to include or exclude neighboring atoms.

FNEU='Y' A single phase-shift for each prototypical neutral atoms is generated.

FNEU='N' Two different mini-cluster are generated in the phagen input for neutral atoms in the case they are surrounded by photoabsorbers or not.

#### 6. NRCD

Number of different maximum neighbor distance cards 7, NRCD $\geq$ 1.

#### 7. SY1 (character\*2) SY2 (character\*2) DDD (real\*4)

Maximum distance DDD (Å) used to define first neighbors between atoms of type SY1 and SY2. The use of wild cards '\*' to indicate all possible atoms is allowed. The information is sequentially updated, for instance the sequence:

'\*' '\*' 2.2  
'Ag' 'I' 3.5

is used to define as first neighbors any couple of atoms at a distance  $\leq 2.2\text{Å}$  or  $\leq 3.5\text{Å}$  in the case of Ag-I bonds.

NOTE: cases b and c (ICM='CRY' or ICM='SYM') crystallographic input only!

#### 7.1.b DCL (REAL\*4)

Cutoff distance (Å) defining the radius around each photoabsorber site of the unit cell reproducing correctly the structure and the number of neighbors. The dimension of the cluster generated by crymol is essentially defined by this parameter.

#### 8. ABC (A3)

format string to identify output file names.



9. NFO, (NF(J), 1, NFO)  
NFO is the total number of inequivalent photoabsorbers for which the calculation of the cluster is required. NF(J) is a vector dimensioned NFO indicating the sequential number of the selected photoabsorbers in the unit cell or molecule.
10. TOLLE  
Tolerance parameter in the fractional distance difference to define inequivalent sites for the phase-shift calculations (suggested values TOLLE=0.05 or 0.10 *i.e.* 5–10%).
11. NAHO  
Type of photoabsorption process: NAHO=1→1s hole, *K*-edge; NAHO=2→2s hole, *L<sub>I</sub>*-edge; NAHO=3→2p<sub>1/2</sub> hole, *L<sub>II</sub>*-edge; NAHO=4→2p<sub>3/2</sub> hole, *L<sub>III</sub>*-edge .....
12. NSR  
Muffin-tin radii selection parameter:  
NSR=0 reads the atomic radii from cards 2.2.a or 2.5.b and uses them in calculation of phase shifts;  
NSR=1 will calculate sphere radii by using the Norman criterion.
13. CIP, EF  
CIP is the core ionization potential (energy unit: Rydberg, 1Ry = 13.606eV) roughly equal to the energy threshold of the absorption edge.  
EF is the supposed Fermi Energy in Rydberg. Not required for complex phase calculation (EF=0.)
14. IMESH, EMIN, EMAX, DE  
IMESH selects between different energy scales in the phase shifts.  
IMESH=1 Constant Energy step (suggested for XANES).  
IMESH=2 Constant wave-vector ( $\propto \sqrt{E}$ ) step.  
IMESH=3 Exponentially increasing Energy step (suggested for EXAFS).  
EMIN First energy point (Rydberg).  
EMAX Last energy point (Rydberg).  
DE Initial step of the energy mesh  
(constant for IMESH=1, variable for IMESH=2, 3)
15. I\_EXC, GAMMA  
I\_EXC exchange and correlation potential model for the excited state:  
I\_EXC=1  $X_\alpha$  Slater approximation  
I\_EXC=2 Dirac-Hara real energy-dependent potential  
I\_EXC=3 Hedin-Lundqvist real energy-dependent potential  
I\_EXC=4 Dirac-Hara complex energy-dependent potential  
I\_EXC=5 Hedin-Lundqvist complex energy-dependent potential  
GAMMA constant imaginary potential (used for options I\_EXC=4, 5) to account for the core hole lifetime in the phase shifts. It represents the Half Width at Half Maximum (HWHM) (Ry). If the input is GAMMA=-1 and the options are I\_EXC=4 or 5, GAMMA is read internally from a table according to M. O. Krause and J. H. Oliver, “*Natural Widths of Atomic K and L levels, K(alpha) X-Ray Lines and Several KLL Auger Lines*”, Phys. Chem. Ref. Data, **Vol. 8 No. 2**, 1979, (p. 329).

## 16. FLS, NSDS, THG, FIG, L1, L2, L3

Selection of the graphical options in the output files `ABC.reg` (REGIS graphic) and `ABC.top` (TOPdrawer graphic) (ABC - see card 8).

FLS='S' selects a stereoscopic plot, otherwise a single plot is produced.

NSDS=1 squint eye vision

NSDS=2 direct vision with lenses of the stereoscopic plot.

THG and FIG specify the rotation angles (degrees) of the image.

L1='S' distinguishes each atom in the cluster with different symbols,

L2='S' labels atoms in the plot,

L3='S' draws the bonds among the atoms of the cluster.

### 3.3. Program run and output files

Under UNIX the program runs with the command

```
crymol < crymol.in
```

where `crymol.in` is the input file written according to the previous description. A detailed diagnostic of the calculations is written on the standard output.

Under VMS, a suitable command file (e.g. `CRYMOL.COM`) is generated adding on top of the the previous input file the appropriate DCL command to run the CRYMOL program (see also Appendix A3).

The program generates several output files:

- NOME.cel contains information on the elementary cell and on the the type and position of the first neighbors identified by the program. This file is not used by the subsequent programs and is mainly useful for checking the input structure. It is generated for the crystalline cases only.
- ABC.ato Symbols, positions, and neighbor tables for all atoms included in the cluster. The first few lines contain a phase shift index table. This file is the required input for the program `gnpeak` to identify the main *n*-atom configurations.
- ABC.leg List of atom pairs identified as chemical bonds with bond-length and versor components. This file is not used and can be useful as a check of the input.
- ABC.in File suitable as an input for the `crymol` program to calculate all the required phase-shifts.
- ABCxyz.xan List of all atoms in the cluster with position in Cartesian coordinates and phase shift index according to the order in the `ABC.in` file. This file can be included in an input file for the program `xanes` to calculate the total cross-section for the central atom in the cluster.
- ABC.reg REGIS file containing commands for 4 levels of intensity or color (VT 330 or VT 240 Digital graphic REGIS terminals). This graphic output is used for high-quality figures of the cluster geometry.
- ABC.top High quality graphical output of the selected atoms cluster suitable for the plotting program TOPdrawer. Using TOPdrawer it is possible to view the plot on most workstations and also to produce a (color) postscript plot.

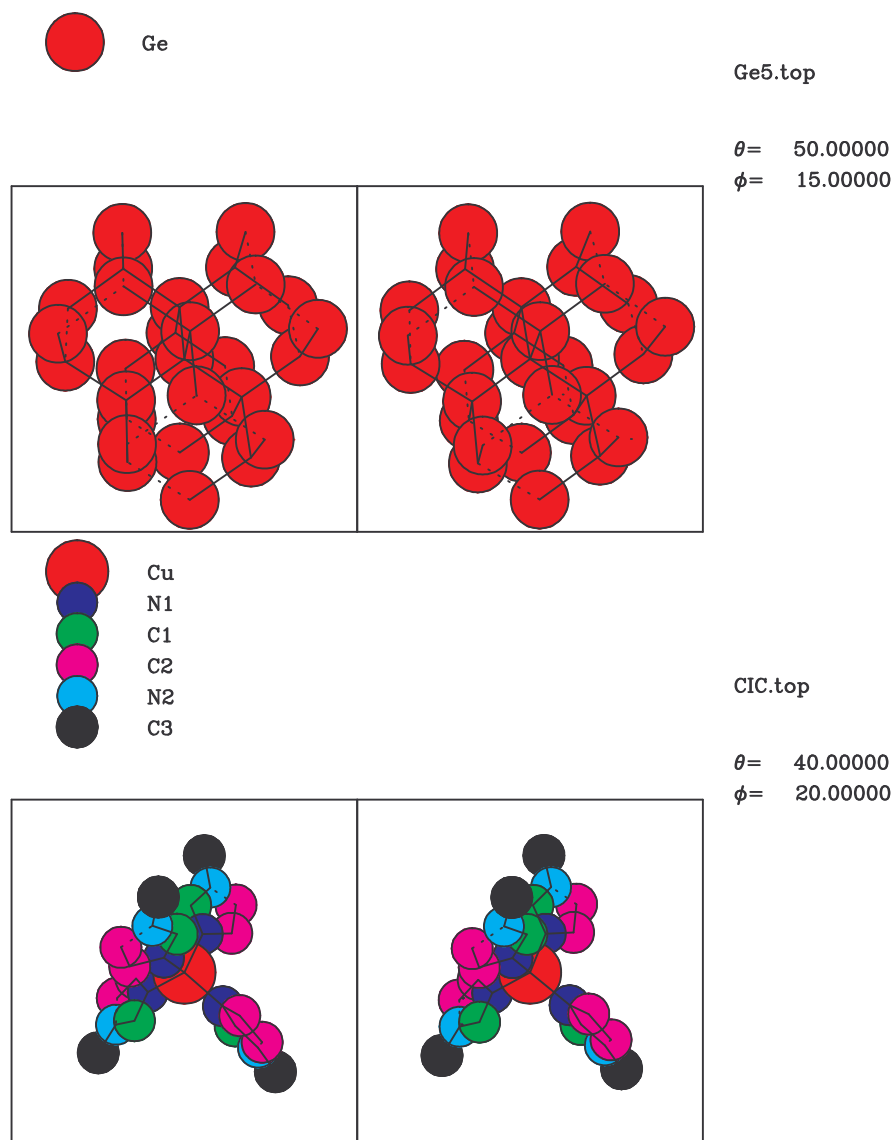


Figure 1. Stereoscopic plots obtained running crymol using the command files a) (c-Ge, upper) and b) (Cu-imidazole, lower) reported in Section 3.4

### 3.4. Examples of command files

a) Input file for cubic c-Ge, with explicit use of the f.c.c. symmetry

```

SYM
c-Ge
C
5.646
2
  1 'Ge'  0.00  0.00  0.00  4  0.0  2.3
  2 'Ge'  0.25  0.25  0.25  4  0.0  2.3
4
0.,0.,0.
    
```

```

0.,0.5,0.5
0.5,0.,0.5
0.5,0.5,0.
1
1,' x      ',' y      ',' z',0.,0.,0.
YNY
1
'*' '*' 2.5
5.
Ge5
1,1
.05
1
1
816.4,0.
3,-.2,100.,0.05
5,-1.
'S',1,50.,15.,'S',',' ','S'

```

## b) Cu-imidazole molecule

```

XYZ
cu4mi
25
1  'Cu'  0.00000  0.00000  0.00000  4  0.0  1.70
2  ' N'  1.56460  0.67739  1.14606  3  0.0  1.40
3  ' N' -1.56460 -0.67739  1.14606  3 -0.0  1.40
4  ' N'  0.67739 -1.56460 -1.14606  3 -0.0  1.40
5  ' N' -0.67739  1.56460 -1.14606  3 -0.0  1.40
6  ' C'  1.53481  1.41954  2.22794  2 -0.0  1.45
7  ' C'  2.88700  0.31732  1.00711  2  0.0  1.45
8  ' C' -1.53481 -1.41954  2.22794  2 -0.0  1.45
9  ' C' -2.88700 -0.31732  1.00711  2  0.0  1.45
10 ' C'  1.41954 -1.53481 -2.22794  2 -0.0  1.45
11 ' C'  0.31732 -2.88700 -1.00711  2  0.0  1.45
12 ' C' -1.41954  1.53481 -2.22794  2 -0.0  1.45
13 ' C' -0.31732  2.88700 -1.00711  2  0.0  1.45
14 ' N'  2.74064  1.53870  2.79171  3 -0.0  1.40
15 ' C'  3.60843  0.83540  2.01620  2  0.0  1.45
16 ' N' -2.74064 -1.53870  2.79171  3 -0.0  1.40
17 ' C' -3.60843 -0.83540  2.01620  2  0.0  1.45
18 ' N'  1.53870 -2.74064 -2.79171  3 -0.0  1.40
19 ' C'  0.83540 -3.60843 -2.01620  2  0.0  1.45
20 ' N' -1.53870  2.74064 -2.79171  3 -0.0  1.40
21 ' C' -0.83540  3.60843 -2.01620  2  0.0  1.45
22 ' C'  3.05408  2.29898  4.00792  1 .000  1.46
23 ' C' -3.05408 -2.29898  4.00792  1 .000  1.46
24 ' C'  2.29898 -3.05408 -4.00792  1 .000  1.46
25 ' C' -2.29898  3.05408 -4.00792  1 .000  1.46
YNY
1
'*' '*' 2.10
CIC
1,1
.10
1
1
660.,0.
3,-.2,30.,.1
5,0.
'S',1,40.,20.,'S',',' ','S'

```

### 3.5. Advanced usage options

The program `crymol` is useful for generating input files for the successive programs. These files are written as text and can be easily edited and modified by the user. In complicated cases a blind program execution is discouraged, since the input files can be, usually, heavily simplified.

The most obvious simplification regards the reduction of required phase shifts. In addition to the action of the flag `FNEU`, small structural differences in the neighborhood of distinct atoms with the same *Z* will generate different phase shifts mini-clusters. The phase differences may be, however, very small in such a way that the atoms differentiation is not justified. In order to reduce the number of phase-shifts the user has to edit the file `ABC.in` to eliminate the corresponding mini-clusters. In order to maintain a record of this phase-shift reduction the phase-shift table in the first lines of the `ABC.at0` file has to be edited and updated accordingly. Let us for example consider the phase shift table for a model structure involving Ag and Br atoms for a Br *K*-edge calculation:

```

3  SIMBOLI
Br  6 VICINI  Z= 35  FASI:   1   2   0
Ag1 8 VICINI  Z= 47  FASI:   0   3   3
Ag2 9 VICINI  Z= 47  FASI:   0   4   4

```

In this case three different symbols (`simboli`) identify Br atoms and two distinct types of Ag atoms differing on the neighbor (`vicini`) table. The three numbers on the right hand side refer to the phase shifts table for the photoabsorber, neutral, and neutral neighbor of photoabsorber cases, respectively. Br atoms are either photoabsorbers (phase 1) or neutral (phase 2). Ag1 atoms are neutral (phase 3). Ag2 atoms use a different phase (4). If, for simplicity, the phase 4 is not calculated in order to use phase 3 for all Ag atoms it will be sufficient to substitute 4 by 3 in the last line

```

Ag2 9 VICINI  Z= 47  FASI:   0   3   3

```

This operation will guarantee that after running `gnpeak` on the corrected `ABC.at0` file the corresponding `ABC.gnx` output, used in the input for `gnxas`, will be consistent with the phase-shifts file generated by `phagen`. The similar substitution of phases 4 by phases 3 should be performed in the `ABCxyz.xan` file prior to perform a `xanes` calculation.

In some molecular systems the automatic construction of the mini-clusters containing the first-neighbors of each prototypical atom is not the best choice for an accurate calculation of phase-shifts by `phagen`. In fact, in several cases, the atoms at the border of the molecule have only one neighbor and this particular situation can affect the muffin-tin radii evaluation, the average interstitial potential and eventually the phase-shifts of those atoms. A simple correction can be made including also the second-neighbors in the `ABC.in` output file defining the mini-clusters for each prototypical atom. For example, for a simple tetrahedral  $\text{SiCl}_4$  molecule a typical `crymol` input file is the following:

```

XYZ
SiCl4
5
1  'Si'  1.16584  1.16584  1.16584  4  0.  1.781656
2  'Cl'  0.00000  0.00000  0.00000  1  0.  2.035534
3  'Cl'  2.33168  2.33168  0.00000  1  0.  2.035534
4  'Cl'  2.33168  0.00000  2.33168  1  0.  2.035534
5  'Cl'  0.00000  2.33168  2.33168  1  0.  2.035534

```

```

Y
1
'Si' 'Cl' 2.5
SC1
1,1
.05
1
0
135.,0.
3,-.2,40.,.1
5,0.0176
'S',1,40.,20.,'S',' ','S'

```

Running `crymo1` would generate the following file `SC1.in` for the successive phagen program:

```

SC1
2
14
2
14 17
135.00000 3 40.00000 -0.20000 0.10000
5 0.01760 0.00000
PHASE CLUSTER N. 1 OF SC1
5 1 1 0 1
' Si' 14 0.000000 0.000000 0.000000 1.781656 0.000000 0.720000
' Cl' 17 -2.203856 -2.203856 -2.203856 2.035534 0.000000 0.720000
' Cl' 17 2.203856 2.203856 -2.203856 2.035534 0.000000 0.720000
' Cl' 17 2.203856 -2.203856 2.203856 2.035534 0.000000 0.720000
' Cl' 17 -2.203856 2.203856 2.203856 2.035534 0.000000 0.720000
PHASE CLUSTER N. 2 OF SC1
2 2 1 0 1
' Cl' 17 0.000000 0.000000 0.000000 2.035534 0.000000 0.720000
' Si' 14 2.203856 2.203856 2.203856 1.781656 0.000000 0.720000

```

A more accurate calculation of the phase-shift for neutral Cl can be easily achieved by inserting again the whole molecule in the second mini-cluster. The following 7 lines must substitute the last 4 lines of the `SC1.in` file:

```

PHASE CLUSTER N. 2 OF SC1
5 2 1 0 1
' Cl' 17 -2.203856 -2.203856 -2.203856 2.035534 0.000000 0.720000
' Si' 14 0.000000 0.000000 0.000000 1.781656 0.000000 0.720000
' Cl' 17 2.203856 2.203856 -2.203856 2.035534 0.000000 0.720000
' Cl' 17 2.203856 -2.203856 2.203856 2.035534 0.000000 0.720000
' Cl' 17 -2.203856 2.203856 2.203856 2.035534 0.000000 0.720000

```

In several molecular cases it is quite easy and often preferable to write directly the input file `ABC.in` for `phagen`. The detailed description of the cards contained in the input file is contained in the `phagen` documentation.

## 4. Program phagen

### 4.1. Introduction

The program `phagen` calculates the atomic phase shifts in the muffin-tin approximation suitable for the calculation of the x-ray absorption fine structure. The program was originally developed by C. R. Natoli and T. A. Tyson under the name `MSCALC_POTGEN`. The

input of the program is a set of atomic and geometrical information on one (or more) prototypical atomic environment present in the structure. For each atomic cluster the program computes the spherically averaged potential around each atomic site up to the respective muffin-tin radius and the corresponding interstitial potential level. Exchange and correlation potential contributions can be included using various approximations. The phase shifts, up to high angular momentum, are calculated for each atomic sites and tabulated in appropriate output files. The output files are suitable inputs for the programs `gnxas` and `xanes` of the package to calculate accurate curved-wave multiple scattering signals or total cross-sections. Useful reference material can be found in ref. [4] (muffin-tin model), [5, 6] (multiple-scattering computation), and [7] (exchange-correlation potentials).

The origin of some of the routines used is indicated below:

- A. The routines XAINPOT and XASYMFN by M. R. COOK and D. A. CASE were modified and used to obtain the symmetrized coordinates and to construct the muffin-tin potentials, respectively. These programs are revised versions of the original multiple scattering programs and are available through the Quantum Chemistry Program Exchange (Chemistry Building Room 204, Indiana University, Bloomington, Indiana 47405, USA).
- B. The routine ATOM\_SUB is the modified routine ATOM taken from the program FEFF [8]. This routine is a revised version of the original atomic relativistic Hartree-Fock-Slater code of J. Desclaux [9].
- C. The routine CONT\_SUB is the continuum cross-section program of C. R. Natoli and D. K. Misemer specialized to the computation of  $T_i$  (related to the scattering phase shift). In this routine an accurate and fast approximation to the real part of the Hedin-Lundqvist exchange potential [6, 8] has been implemented.

#### 4.2. Input command file

The input files for the program `phagen` can be automatically generated using the `crymol` program, however, for simple cases they can be written directly or edited by the user (see `crymol` documentation). Examples of a simple molecule  $\text{BBr}_3$  with the card description is reported below. An input file for the  $\text{SiCl}_4$  molecule is reported in the `crymol` documentation.

The general idea behind the program is that the phase-shifts have to be calculated, at least, for every atomic species present in the structure. The phase shifts will depend primarily on the atomic number  $Z$ , however, they will be also different for a neutral or for the photoabsorber atom since the core-hole and the resulting charge relaxation will modify slightly the potential. Finally, there will be a small “chemical” difference according to the immediate environment of the atoms. This chemical effect is accounted for by the charge overlap from the neighboring atoms. The user will select the number of different phase shift to consider for photoabsorber and backscattering atoms. For each phase-shift the program will require information on the central atom (possible core-hole presence) and on the type and position of surrounding atoms. This set of atoms will be referred to as “mini-cluster”. For each inequivalent site in the molecule or crystal the environment will be in principle different, however, the phase-shift difference will be usually negligible, for the same  $Z$ , so that a single average structure for every  $Z$  is usually sufficient. In the case of a glass or a

liquid where the atomic neighborhood is only defined in an average sense a single phase-shift calculated for a mini-cluster reproducing the average neighbors with an appropriate symmetry will be sufficient.

The size of the mini-clusters should be wisely chosen. For a simple molecule it is possible to include always all of the atoms (see  $\text{SiCl}_4$  and  $\text{BBr}_3$  cases). This choice guarantees the the phase shifts for the different atoms are exactly referred to the same interstitial potential level. In the case of complicated structures the limited size of the mini-cluster introduces approximations in the phase-shift calculation. In the majority of cases, however, the inclusion of the first shell environment is sufficient.

The input file contains 7 cards which provide the calculation parameters followed by a set of cards for each prototypical phase to be calculated. The phaseshift will refer to the first atoms in each mini-cluster list.

*Important:* the units in the input file prepared by `crymo1` are currently atomic units (a.u., Rydberg). It is possible to use also practical units (eV, Å). The actual choice is selected in card 5 by parameter IMESH. If it is negative (eV, Å) units are used. If it is positive (default choice) (a.u., Rydberg) are used (Conversion factors 1 Ryd = 13.61 eV, 1 a.u. = 0.529 Å).

#### 4.2.1. Detailed card description

CARD 0: String to identify the output file names, (A3) format.

CARD 1: Number of phase-shifts to be calculated.

CARD 2: Atomic number of the photoabsorber.

CARD 3: Number of different neutral atomic numbers involved in the calculation.

CARD 4: Atomic numbers of the various neutral atoms involved. Must be as many as specified by card 3.

CARD 5: CIP, IMESH, EMAX, EMIN, DE (final state)

CIP core ionization potential used to compute  $\hbar\omega$  in  $\sigma_0$  (see for example the *electron binding energies* in the X-ray Data Booklet *J. Kirz et al.*, p. 2-6 to 2-11, 1986, [http://www-cxro.lbl.gov/data\\_booklet](http://www-cxro.lbl.gov/data_booklet));

IMESH energy mesh type:

1 = constant energy step (useful for XANES),

2 = constant  $k$  step (useful for EXAFS),

3 = exponential mesh versus energy (useful for EXAFS);

*a negative sign means all units are (eV, Å)*

EMAX energy maximum;

EMIN energy minimum;

DE energy spacing (initial if IMESH=3);

CARD 6: I\_EXC\_POT, GAMMA, FERMI\_ENERGY

I\_EXC\_POT Exchange and correlation potential type (final state)

1 = X-alpha exchange

2 = Real Dirac-Hara exchange

3 = Real Hedin-Lundqvist exchange

4 = Complex Dirac-Hara exchange

5 = Complex Hedin-Lundqvist exchange



GAMMA Constant imaginary potential (for I\_EXC\_POT = 4 or 5)  
(HWHM for core-hole)

FERMI\_ENERGY Energy of the Fermi level (usually 0.)

CARD 7: Comment header describing mini-cluster

CARD 8: NAT, I\_ABSORBER, I\_ABSORBER\_HOLE, I\_NORMAN, I\_MODE

NAT number of atoms in the mini-cluster,

I\_ABSORBER the next I\_ABSORBER<sup>th</sup> atom has a core hole (photoabsorber). If '0' all atoms are neutral,

I\_ABSORBER\_HOLE type of core-hole present in the I\_ABSORBER<sup>th</sup> atom.

1 = (1s) *K*-edge

2 = (2s) *L<sub>I</sub>*-edge

3 = (2p<sub>1/2</sub>) *L<sub>II</sub>*-edge

4 = (2p<sub>3/2</sub>) *L<sub>III</sub>*-edge

5 = *M<sub>I</sub>*-edge

6 = *M<sub>II</sub>*-edge

7 = *M<sub>III</sub>*-edge

8 = *M<sub>IV</sub>*-edge

9 = *M<sub>V</sub>*-edge

I\_NORMAN select algorithm for defining muffin-tin radii

0 = calculate potential with user supplied radii

1 = use scaled Norman radii (see REDF, card 9)

I\_MODE final potential type parameter (not yet implemented)

0 = generate potential internally

1 = read in external potential

CARD 9: SYM, NZ, XV, YV, ZV, RS, CHARGE, REDF

NAT cards of this type specify data for each atom. The phase shifts are calculated only for the first atom in the list. At least one neighbor has to be included. The sequence of cards 7, 8, and 9 must be repeated for each mini-cluster.

SYM atomic symbol, (A5) format,

NZ atomic number (Z),

XV *x*-coordinate,

YV *y*-coordinate,

ZV *z*-coordinate,

RS muffin-tin radius (used if I\_NORMAN=0)

CHARGE local charge (not implemented),

REDF RS=R\_NORMAN\*REDF (if I\_NORMAN=1) reasonable values are in the 0.7–0.8 range.

### 4.3. Examples of input file

Input file for the BBr<sub>3</sub> molecule using (eV, Å) units. Phase shifts calculated for Br (photoabsorber), Br (neutral), and B, respectively.

```

BBr
      3
     35
      2

```

```

      35      5
13480.0 -2 2448. -5. 0.68
      5      1.36      0.0
PHASE CLUSTER N. 1 OF BBr3
      4      1      1      0      1
' Br' 35 -1.881      0.0      0.0      1.1701      0.0      0.72
' B'  5  0.0000      0.0      0.0      0.8330      0.0      0.72
' Br' 35  0.9403      1.6286      0.0      1.1701      0.0      0.72
' Br' 35  0.9403     -1.6286      0.0      1.1701      0.0      0.72
PHASE CLUSTER N. 2 OF BBr3
      4      3      1      0      1
' Br' 35 -1.881      0.0      0.0      1.1701      0.0      0.72
' B'  5  0.0000      0.0      0.0      0.8330      0.0      0.72
' Br' 35  0.9403      1.6286      0.0      1.1701      0.0      0.72
' Br' 35  0.9403     -1.6286      0.0      1.1701      0.0      0.72
PHASE CLUSTER N. 3 OF BBr3
      4      2      1      0      1
' B'  5  0.0000      0.0      0.0      0.8330      0.0      0.72
' Br' 35 -1.881      0.0      0.0      1.1701      0.0      0.72
' Br' 35  0.9403      1.6286      0.0      1.1701      0.0      0.72
' Br' 35  0.9403     -1.6286      0.0      1.1701      0.0      0.72

```

#### 4.4. Program run and output files

Let `phagen.in` be the appropriate input file, then `phagen` runs with the simple UNIX command:

```
phagen < phagen.in
```

or running a suitable command file under OpenVMS (see Appendix A3).

The program generates 5 output files identified by the (A3) string ABC given in the input file. Their name and content are listed below:

```

ABCpha.chk log file containing information on symmetry and phase-shifts calculation;
ABCpha.out log file containing information on the muffin-tin radii determination;
ABCpha.pot tabulated potentials;
ABCpha.s0 atomic cross-section file;
ABCpha.t1 phase-shifts file ( $t_\ell$  matrices).

```

Only the last two files are required as input for the programs `gnxas` or `xanes`.

Muffin-tin radii and interstitial potential values are also written directly on the screen. The phase-shifts can be viewed with the help of the program `phaplo` which reads the `ABCpha.t1` file (supplied as a command argument in UNIX)

```
phaplo ABCpha.t1
```

and produces TOPdrawer plot files for each atomic phase shift.

#### 4.5. Guidelines for the choice of muffin-tin radii

Muffin-tin (MT) radii should be regarded as empirical calculation parameters. The final calculated x-ray absorption fine structure signal will be weakly dependent on the actual radii provided that they are chosen in a reasonable range. The covalent atomic radii can be used as a first guess MT radii. For a homoatomic covalent bond tangent radii of equal size represent a reasonable choice. As an example in a  $\text{Br}_2$  molecule ( $R = 2.29 \text{ \AA}$ ) the optimal choice is about  $R_{\text{MT}} = 1.14 \text{ \AA}$ . For the Ge crystal where  $R \approx 2.45 \text{ \AA}$ ,  $R_{\text{MT}} = 1.22 \text{ \AA}$  etc. In the case of a metallic system short MT radii down to 0.7–0.8 of the tangent condition

can be used. The advantage of using short MT radii is in the smaller number of angular momenta required to describe the atomic scattering. The maximum angular momentum is in fact of the order of  $l_{\max} = k_{\max} R_{\text{MT}}$ . A too small MT radius, however, is likely to leave too much of the electronic charge in the interstitial region and should be avoided, especially for covalent bonds. The electronic charge contained in the MT spheres can be easily checked looking at the ABCpha.out output file.

For different Z the ratio among the relative MT radii can be taken equal to the ratio of the corresponding Norman radii. These are defined in such a way that the spherically integrated electronic charge, including the charge of the neighboring atoms up to a distance  $R_N$  equals Z. Norman radii are calculated by phagen and if the flag I\_NORMAN=1 is used, these radii, scaled by the factor REDF, are used as MT radii. Typical values of this scaling factor are in the range 0.7–0.8 (current value set by crymo1 0.72). Notice that  $R_N$  is calculated correctly only if all the neighbors of the atom are specified. For mini-clusters containing only the first shell of neighbors  $R_N$  and  $R_{\text{MT}}$  will be estimated correctly only for the first (central) atom in the list. As a consequence phagen should run twice for these cases, one with I\_NORMAN=1 to determine the  $R_{\text{MT}}$  for all types of atoms and a second time with I\_NORMAN=0 after writing in all the card 9 input lines the corresponding previously determined  $R_{\text{MT}}$ , as the input variable RS.

## 5. Program gnpeak

### 5.1. General program description

The program gnpeak is a utility program able to analyze an input atomic cluster and to identify the main 2, 3, and 4 atoms sub-clusters relevant for the calculation of the leading single and multiple scattering contributions to the x-ray absorption cross-section. The input for the program can be generated automatically using crymo1, some of the output files can be used directly as part of the input for the signal calculation program gnexas.

The program identifies all the distinct two, three, and four-body local configurations around a photoabsorber atom. These atomic configurations will be often referred to as “peaks” of the two-body ( $g_2$ ), three-body ( $g_3$ ), and four-body ( $g_4$ ) distribution functions, respectively. The nomenclature is borrowed from the theory of the liquid structure. In a crystalline or molecular case the  $n$ -body distribution functions ( $g_n$ ) are the sum of thermally broadened peaks centered around equilibrium geometries. Each configuration generates several single and multiple scattering contributions to the x-ray absorption cross-section.

The analysis is performed on a cluster, specified by an input formatted file denominated ABC.ato (ABC is a three character string), that contains basic information on the atoms including type, position, and neighbours. The file can be prepared automatically by the program crymo1 (see crymo1 documentation for details). The cluster can be either representative of the whole structure (as in the case of a molecule) or of a portion of it up to a distance D from the central photoabsorber atom, as in the case of a crystal. In this latter case the sub cluster analysis can be performed up to maximum distance  $\text{RCUT} \leq D$  for which the counting of the atomic configurations is still correct in the cluster. If the user requires  $\text{RCUT} > D$  a new crymo1 run with the appropriate input should be executed. For each photoabsorber atom the program selects all possible different arrangements of 1, 2 and 3 neighbouring atoms. The output is a table of peaks of the pair, triplet and quadruplet distribution functions  $g_2$ ,  $g_3$  and  $g_4$  with the constraint that one of the atoms must be of the

photoabsorber type. Possible equivalence of peaks after a permutation of the photoabsorber are considered. This can reduce considerably the number of different structures that have to be considered in a case in which many atoms of the photoabsorber type exist in the structure.

### 5.2. Input files

The only file required is ABC.ato where ABC can be any (A3) string. ABC.ato can be automatically generated by the crymol program.

### 5.3. Program execution

The program execution is trivial and suitable for an interactive run.

CARD 1: ABC (character\*3) → cluster name ... will read file ABC.ato

CARD 2: EPS (free format) → tolerance distance in Å within which two geometries are considered equivalent. A simple return or zero can be typed for a default option of .0002 Å.

CARD 3: RCUT (free format) → cutoff distance for the sub-cluster size counting. The value zero or just return can be entered to select the maximum size compatible with the cluster.

NOTE: *in case the number of atoms exceeds 100 an additional card is needed.*

CARD 3a: RCUT4 (free format) → cutoff distance for the paths in the  $g_4$ , may be selected smaller than RCUT to limit the number selected 4-atom clusters.

### 5.4. Output files

The program generates four output files:

→ ABC.gnp

→ ABC.gnx

→ ABC.chi

→ ABC.dbw

ABC is the same (A3) string of the input file, the extensions are given by the program automatically. The open qualifier is STATUS='UNKNOWN', so that previous version are overwritten under UNIX operating systems. The file ABC.gnp contains a list of the  $g_2$ ,  $g_3$ , and  $g_4$  peaks found in the cluster below the given cutoff, the format is useful for a rapid inspection of the peak analysis performed by gnpeak. The file ABC.gnx contains the same information of the ABC.gnp file, but in numerical form directly usable as part of an input for the gnxtas program which actually performs the signal computations (see the gnxtas documentation for details). The file ABC.chi contains a list of the main  $\chi_n$  paths related to the  $g_n$  peaks. First the  $\chi_2$  and  $\chi_4$  related to a  $g_2$  distribution are listed, then all the  $\chi_3$ , and  $\chi_4$  related to a  $g_3$ , and finally all the  $\chi_4$  related to  $g_4$  peaks are listed. The association of  $\chi_n$  contribution to  $g_n$  peaks actually follows a physical intuition, since all the paths generated by the same geometrical arrangement of 2, 3 or 4 bodies are in this way listed together. The use of this file is mainly to provide a link with previous approaches where a wide use of the multiple scattering series was performed. Finally the file ABC.dbw is generated only in the case when the cluster is disordered and it has been analysed with a large EPS tolerance so that several configurations are included into the same peak. It

contains the structural covariance matrices of the geometrical parameters for the various  $g_2$  and  $g_3$  peaks. Covariance matrices are not calculated for  $g_4$  peaks.

### 5.5. Peak counting scheme

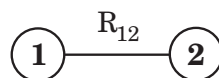
The counting of the relevant  $n$ -body configurations (sub-clusters) in a given structure, involves the evaluation of equivalent atomic arrangements that, if suitably rotated and translated, can be overlapped atom by atom (coincidence in position and atomic types). The number of equivalent atomic configurations divided by the number of atoms in the structure is referred to as PEAK DEGENERACY. As an example in a f.c.c. lattice there are 6 two-body configurations, with two identical atoms placed at the first neighbour distance  $R$ , per atom in the crystal. The degeneracy of this two-body configuration is therefore 6.

There is however a different “local” perspective for the configuration counting which considers what is seen from a given atom in the structure. This is exactly what is required for calculating the x-ray absorption signal due to a photoabsorption process involving that atomic site. This counting involves the conditional  $n$ -body distribution functions that describe the probability that  $n$  atoms are placed in a given arrangement provided that one of them is of a specified type and is in a given position. The two configuration counting schemes are equivalent. Coming back to the previous example, in the f.c.c. lattice is well known that the number of first neighbours at the distance  $R$  is 12. This is the so called COORDINATION NUMBER corresponding to the “first peak” of the “conditional two-body distribution function”. Thus the two-body configuration with a degeneracy of 6 generates a shell of neighbours with coordination number of 12. The doubling from 6 to 12 arises because both of the atoms in the two-body configuration are identical and can act as the photoabsorber. More complex relationships occur in the case of three and four-body peaks.

Let us now establish what are the parameters sufficient to describe a  $n$ -body peak, starting with the two-body ones.

#### 5.5.1. Two-body peaks

A two-body peak, that can be schematically represented as:



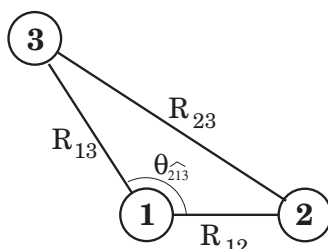
is completely described by the specification of the distance and of the two atomic types.

$R_{12}$ ,  $TY_1$ ,  $TY_2$

Only those peaks in which at least one of the two atoms is of the photoabsorber type atoms are interesting for the EXAFS calculation. Let  $A_2$  be the degeneracy of the peak. Then there are two possible cases for the counting in the local perspective. Either  $TY_1$  is different from  $TY_2$  and then the contribution to the number of configuration (conf #)  $NC_2$  (the so called “coordination number”) is  $NC_2=A_2$  or  $TY_1=TY_2$  and  $NC_2=2 \times A_2$  (as in the previous example).

### 5.5.2. Three-body peaks

A three-body peak can be schematically represented as:



It is completely described by three roto-translational invariants and the atomic types in positions 1, 2 and 3. In light of the chemical applications of the theory where the most stable and well defined distances are likely to be the shortest two among the three sides of the triangle and the chemically meaningful bond angle be the angle between these two sides, the following convention will be used:

- 1) the peak geometry is established by the distances of the shortest bonds  $R_{12}$  and  $R_{13}$  and the angle  $\theta_{213}$  in between;
- 2) the site 1 is by definition the vertex between the two shortest bonds;
- 3) the site 2 is by definition the vertex at the other end of the shortest bond;
- 4) the site 3 is the other site in the triangle.

If the application of these does not define a unique choice, then any choice is completely equivalent for symmetry reasons. Summarizing, the peak position is specified by the parameters  $R_{12}$ ,  $R_{13}$  and  $\theta_{213}$  ( $R_{12}$ ,  $R_{13}$ ,  $\theta_{213}$ ) and by the atomic types of the atoms in position 1, 2 and 3 respectively  $TY_1$ ,  $TY_2$  and  $TY_3$ ;  $R_{23}$  can be calculated easily by using the Carnot formula. The degeneracy of the peak will be indicated with  $A_3$ .

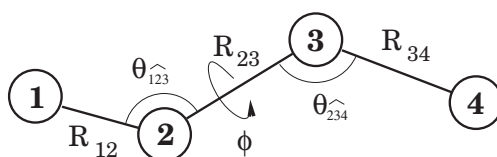
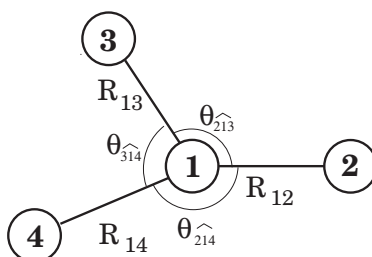
There are three possible symmetry conditions on the triangle which can be checked with a permutation of the atoms: in the most general case when the three sides, or the atoms, are not equivalent the triangle is scalene "S". If two atoms are equivalent as well as the two bond connecting them to the third atom, then the triangle is isosceles "I". Finally if both three atoms and sides are the same ( $\theta = 60^\circ$ ) the triangle is equilateral "E". Now it is important to consider the various possible positions of the photoabsorber since each different position will generate a different structural signal. Similarly to the neighbour coordination number, each of these conditional three-body arrangements will be associated with a coordination number, counting the number of configurations of this kind seen from a given photoabsorber. Clearly each of the three possible positions of the photoabsorber will generate  $A_3$  contribution to a coordination number.

According to the symmetry properties sometimes the number of configurations (conf #)  $NC_3$  can be  $2 \times A_3$  or  $3 \times A_3$  (equilateral case).

### 5.5.3. Four-body peaks

In a four-body peak there are six inter-atomic distances, the geometry is completely described by six roto-translational invariants and the atomic types in positions 1, 2, 3, and 4.

In light of the chemical applications of the theory, where the most stable and well defined distances are likely to be the shortest ones among the six sides of the tetragon



and the chemically meaningful bond angle be the angle between these sides, the following convention will be used:

- 1) the peak geometry is established by a choice of three bonds whose distances sum up to the smaller possible value, these distances can be connected either in a “star” way or in a “chain” way. The peak Si, Cl, Cl, Cl in a SiCl<sub>4</sub> molecule for instance has a natural “star” topology with all meaningful bonds (Si–Cl) irradiating from the Si atom. The peak Br–C–C–C of the molecule Bromo-propane has a natural “chain” topology with the sequence of consecutive bonds Br–C, C–C, and C–C;
- 2) if a star topology occurs then it is parameterized using the following distances and angles in the given sequence:

$$R_{12}, R_{13}, \theta_{213}, R_{14}, \theta_{214}, \theta_{314}$$

and the atom types are

$$TY1, TY2, TY3, TY4$$

Notice that position 1 is the central position in the star, position 2 is at the end of the first distance  $R_{12}$ , position 3 is at the end of the second distance  $R_{13}$ , and position 4 is at the end of the third distance  $R_{14}$ . The three angles  $\theta_{213}$ ,  $\theta_{214}$  and  $\theta_{314}$  are three angles around the vertex 1;

- 3) if a chain topology occurs then it is parameterized by using the following distances and angles in the given sequence:

$$R_{12}, R_{23}, \theta_{123}, R_{34}, \theta_{234}, \phi$$

and the atom types are

$$TY1, TY2, TY3, TY4$$

Notice that position 1 is the first end of the chain, position 2 is at the other end of the first distance  $R_{12}$ , position 3 is at the end of the second distance  $R_{23}$ , and position 4 is at the other end of the chain with respect to position 1, at the end of the third distance  $R_{34}$ . The angle  $\theta_{123}$  is the one between the first two bonds, the angle  $\theta_{234}$  is the one between the second and third bonds of the chain. The actual spatial orientation of the three bonds is finally defined by the corresponding *Dihedral angle*  $\phi$ .

The degeneracy of the peak will be indicated by

A4

Notice that similarly to the  $g_3$  case the same four-body peaks can generate several (up to four) different conditional peaks with the photoabsorber placed in different positions. And each permutation of the photoabsorber can be either singly degenerate or two-fold, three-fold and also four-fold degenerate. This degeneracy times A4 is indicated by the number of configurations (conf #) NC4.

### 5.6. Details on the ABC.gnp file

- $g_2$  peaks  $\rightarrow$  I, R12, NC2, TY1, TY2
  - I progressive number 1,2,3,... of the peak
  - R12 distance in Å
  - NC2 number of atoms
  - TY types of the atoms in position 1 and 2
- $g_3$  peaks  $\rightarrow$  I/J, R12, R13, TH, NC3, PhP, TY1, TY2, TY3, NA3(3), FR3
  - I progressive number 1,2,3,... of the peak
  - J equivalence label: all peaks with the same J belong to the same geometry. They differ in the photoabsorber position.
  - R12 bond length between atoms 1 and 2 in Å
  - R13 bond length between atoms 1 and 3 in Å
  - Th bond angle  $\theta_{213}$
  - NC3 number of triplets for each photoabsorber
  - PhP photoabsorber position 1, 2, or 3
  - TYn types of the three atoms involved
  - NA3 the order number of the three atoms in the cluster
  - FR3 peak leading frequency in Fourier spectrum

NOTE: An additional line, if present, indicates that the third and longer bond coincides with a previously classified two-body configuration. This will be useful to reduce the number of independent coordinates in the system.

- $g_4$  peaks:
  - a) if FT='S'  $\rightarrow$  I/J, FT, R12, R13, T213, R14, T214, T314, NC4, PhP, TY1, TY2, TY3, TY4, NA4(4), FR4
  - b) if FT='C'  $\rightarrow$  I/J, FT, R12, R23, T123, R34, T234, DihA, NC4, PhP, TY1, TY2, TY3, TY4, NA4(4), FR4
    - I progressive number 1,2,3 ...of the peak
    - J equivalence label: all peaks with the same J belong to the same geometry. They differ in the photoabsorber position.
    - FT configuration type 'S'= star, 'C'= chain
    - R12 bond length  $R_{12}$  between atoms 1 and 2 in Å
    - R13 bond length  $R_{13}$  between atoms 1 and 3 in Å
    - R23 bond length  $R_{23}$  between atoms 2 and 3 in Å
    - R14 bond length  $R_{14}$  between atoms 1 and 4 in Å
    - R34 bond length  $R_{34}$  between atoms 3 and 4 in Å
    - T213 bond angle  $\theta_{213}$



T214 bond angle  $\theta_{\widehat{214}}$   
 T314 bond angle  $\theta_{\widehat{314}}$   
 T123 bond angle  $\theta_{\widehat{123}}$   
 T234 bond angle  $\theta_{\widehat{234}}$   
 DIHA Dihedral angle  $\phi$   
 NC4 number of quadruplets for each photoabsorber  
 PhP photoabsorber position 1, 2, 3, or 3  
 TYn types of the four atoms involved  
 NA4 the order number of the four atoms in the cluster  
 FR4 peak leading frequency in Fourier spectrum

### 5.7. Details on the ABC.chi output file

This output file contains the information on the:

- associated  $\chi_2$  and  $\chi_4$  paths  $\rightarrow$  I, Xn, PATH, SYP, RP, DEG, NP
  - I progressive number 1,2,3,... of the  $\chi_2$  or  $\chi_4$  contributions  $\chi_n$  is  $\chi_2$  or  $\chi_4$  according to the symbol
  - PATH a sequence of 3 or 5 atoms number indicating a prototypical path in the cluster
  - SYP a sequence of 3 or 5 symbols indicating atoms type involved
  - RP path length \*0.5, *i.e.* leading frequency in Fourier spectrum
  - DEG degeneracy of the path
  - NP number of  $g_2$  peak which generates the path
- associated  $\chi_3$  and  $\chi_4$  paths  $\rightarrow$  I, Xn, PATH, SYP, RP, DEG, NP, NPE
  - I progressive number 1,2,3,... of the  $\chi_3$  or  $\chi_4$  contributions  $\chi_n$  is  $\chi_3$  or  $\chi_4$  according to the symbol
  - PATH a sequence of 4 or 5 atoms number indicating a prototypical path in the cluster
  - SYP a sequence of 4 or 5 symbols indicating atoms type involved
  - RP path length \*0.5, *i.e.* leading frequency in Fourier spectrum
  - DEG degeneracy of the path
  - NP progressive number 1,2,3,... of the peak
  - NPE equivalence label: all peaks with the same J belong to the same geometry. They differ in the photoabsorber position.
- associated  $\chi_4$  paths  $\rightarrow$  I, Xn, PATH, SYP, RP, DEG, NP, NPE
  - I progressive number 1,2,3,... of the  $\chi_4$  contributions  $\chi_n$  always  $\chi_4$  in this case
  - PATH a sequence of 5 atoms number indicating a prototypical path in the cluster
  - SYP a sequence of 5 symbols indicating atoms type involved
  - RP path length \*0.5, *i.e.* leading frequency in Fourier spectrum
  - DEG degeneracy of the path
  - NP progressive number 1,2,3,... of the peak
  - NPE equivalence label: all peaks with the same J belong to the same geometry. They differ in the photoabsorber position.

This output provides a clear table to identify the contribution of interest in the specific case.

## 6. Program gn<sub>x</sub>as

### 6.1. General description

The program gn<sub>x</sub>as is an advanced versatile code for the calculation of x-ray absorption signals associated with (up to 4) atom clusters. The acronym is composed of gn, indicating the  $n$ -body distribution functions ( $g_n$ ) in disordered matter, and xas, X-ray Absorption Spectroscopy. The name reflects the underlying philosophy connecting XAS signals with precise structural features in condensed matter and the possibility to tackle even liquid phase spectra. The program is able to calculate the signals associated with 2, 3, and 4 atoms clusters representing reference configurations in the two-body  $g_2$ , three-body  $g_3$  and four-body  $g_4$  distribution functions. Several algorithms are implemented to this purpose. The input of the program requires a phase-shift and atomic cross-sections ( $\sigma_0$ ) files, generated by the phagen program, and appropriate geometrical information, that can be input either by hand or generated automatically using the gnpeak program. The main output of the program is made of binary files, for each calculated signal, containing amplitude and phase functions vs. energy (or  $k$ ) together with their derivatives with respect to all geometrical coordinates. This information is essential to be able to perform a fitting of the average coordinates and disorder parameters. These binary files (extension .der) are in fact the main input for the successive fittheo program. The program also generates several plot files for viewing all calculated signals.

### 6.2. Introduction

#### 6.2.1. Distribution Functions

In any type of structure (molecule, crystal, liquid, amorphous solid, ...) it is possible to describe the average local environment around the photoabsorber atomic species in terms of  $n$ -body distributions. The radial distribution function  $g_2(r)$  represent the ratio between the local density of surrounding atoms with respect to the average density. The average number of atoms surrounding the photoabsorber will be given by the integral of the  $4\pi\rho r^2 g_2(r)$  function. These definitions can be generalized to the case of multi-component systems. For molecular or crystalline structures the  $g_2(r)$  is given by the sum of a series of peaks associated with specific configurations. The effect of thermal broadening in the low-temperature harmonic limit is to introduce a Gaussian broadening. The variances of these distributions are usually referred to as Mean Square Relative Displacement (MSRD). At very low temperatures the MSRD will be determined by the zero point quantum vibrations, while, at larger temperatures, the MSRD will follow the classical behaviour increasing proportionally to the temperature. At high temperatures (sometimes before the classical regime starts) anharmonic effects may be important.

In the case of amorphous solids, the disorder in the equilibrium positions of the atoms produces already broad distributions that are subject to additional thermal broadening effects. In a liquid, where diffusion takes place, structural and thermal broadening are not distinguishable anymore. In both these latter cases the radial distribution function assumes a smooth shape and is usually characterised by a continuous distribution of atomic distances. A sharp first neighbor peak is usually present in these system whereas in the long range limit the  $g_2(r)$  tends to 1 due to the lack of correlation between distant atoms positions.

The  $g_3$  and the  $g_4$  describe, in a similar way, the probability distribution of 3 and 4 atoms cluster. They contain therefore a more detailed information than the  $g_2(r)$ . As for the  $g_2(r)$ , in the case of molecules or crystals, the  $g_3$  and  $g_4$  can be modeled by a series of Gaussian peaks. In the case of liquids or amorphous solids they are composed of a series of broad peaks at short range which merge into an uncorrelated uniform region in the long range limit.

### 6.2.2. Signal Calculation

A very important feature of the  $g_n$  distribution functions is that they are usually composed of well defined peaks in the short range region. The reason underlying this rather general statement is in the existence of more or less well defined bonds between the atoms. The contribution to the XAS signal from an isolated peak can be calculated from the undamped signal corresponding to the central peak position with small amplitude and phase corrective terms to account for the peak broadening.

A typical example of these corrections is the so called EXAFS Debye-Waller factor  $\exp(-2\sigma^2k^2)$  accounting for the Gaussian damping of the Single Scattering EXAFS terms. The general theory of the configurational damping of the signal associated with a given  $n$ -body configuration or Multiple Scattering (MS) path was reported elsewhere [10]. The principle of this approach on which our signal calculation scheme is based, suggests that in order to perform a configurational average over a relatively narrow distribution it is important to know the signal shape on slightly displaced geometries (also useful to calculate the effect of a shift of the central position). This can be achieved through the calculation on a single reference signal at a central position, together with the coefficients of low order Taylor expansions of the signal amplitude and phase with respect to the geometrical parameters. These calculations for reference positions are performed by the `gnxas` program.

The XAFS signal associated with a continuous distributions of atom as specified by a  $g_2(r)$  function can be instead calculated using the program `grxas`. The reader should refer to the `grxas` documentation for details (Section 9.3).

### 6.2.3. Independent Coordinates – Permutations

The structural units of interest for the XAFS signal calculations are usually all the near neighbor typical bond distances, or short-range three or four-atoms sub-clusters. There is a hierarchical relation between such atomic configurations, indeed a three-body configuration contains two different typical bond lengths involving the photoabsorber. Conversely the same two-body configuration will be contained in several three-body configurations. Similar relationships involve four-body and three-body configurations. For the above reasons the number of independent geometrical parameters that describe the short range of a given structure is usually limited in spite of the large number of  $g_2$ ,  $g_3$ , and  $g_4$  peaks. A noticeable case of redundancy occurs in the cases of systems with many equivalent photoabsorber-type atoms, and consequently in the limit case of a pure elemental system.

Let us take a triangle of equivalent atoms. This structure will generate three different signals according to photoabsorber position. The three signals are not independent and they change simultaneously as the atoms coordinates move. From the point of view of the signal calculations it is very important to use the same coordinates in each calculation involving any of the three possible photoabsorber positions.

In the case of triangles with a particular symmetry, isosceles or equilateral, the number of independent permutations of the photoabsorber is reduced, to two and one, respectively.

A similar occurrence should be considered for the four-body configurations. The relevant reference geometries in a complex molecular or crystalline structure can be identified using the program `gnpeak`.

#### 6.2.4. $\chi$ , $\gamma$ , and $\eta$ signals

The next step in the analysis consists in the calculation of the various signal contributions associated with these reference geometries. Each  $n$ -atom configuration contributes with one (or more, for every inequivalent photoabsorber position) irreducible  $n$ -body XAS signal denominated  $\gamma^{(n)}$ . These are defined as the contribution to the EXAFS signals due to the simultaneous presence of the  $n-1$ -atom cluster surrounding the photoabsorber [1].

Two computational strategies are available for calculating these signals: they can be calculated directly using fast algorithms based on the evaluation of continued fractions [11] or alternatively they can be expanded into the corresponding MS series. As an example the  $\gamma^{(2)}$  signal can be expanded in a MS series involving the successive scattering on the potential of the photoabsorber and neighboring atoms, pictorially represented as:

$$\gamma^{(2)} = \text{---} \circ \text{---} \circ = \text{---} \circ \rightleftarrows \circ + \text{---} \circ \equiv \equiv \circ + \text{---} \circ \equiv \equiv \equiv \circ + \dots$$

Conceptually similar (though slightly more complicated) expansions can be worked out for higher order  $\gamma^{(n)}$  terms [1]. The MS expansion formalism was originally developed for the calculation of the entire absorption cross-section providing the higher order corrections to the first order single scattering signal (see [5,6] and ref. therein).

Both computational strategies present advantages and disadvantages. Generally speaking in the low energy part of the spectrum the continued fraction approach has better convergence properties, whereas, at higher energies, the MS expansion approach is computationally convenient. The main drawbacks of the MS expansion are:

- i) lack of convergence in the low  $k$  region,
- ii) computational difficulties above the first orders, and
- iii) proliferation of paths beyond the  $\chi_3$  and  $\chi_4$  level that make the relationship between structure and signal cumbersome.

In order to avoid signal proliferation in a fitting procedure it is essential to group as many signals as possible provided that they are associated with the same  $n$ -atom configuration. This procedure is convenient especially in the case of leading signals with comparable frequencies in such a way that the total effective signal is characterized by a smooth dependence of amplitude and phase as a function of energy and structural parameters. This is important to validate the Taylor expansion approach for the configurational average calculation [10]. In this case, instead of considering all the MS paths separately, it is preferable to work with the  $\gamma^{(n)}$  signals directly or even to group several  $\gamma^{(n)}$  signals together when applicable. Conversely, if the smoothness condition is not achieved it is better to treat the terms separately. A noticeable exception to the grouping advantage occurs sometimes for the  $\gamma^{(2)}$  term, where the first MS correction  $\chi_4$  (to the  $\chi_2$ ) has the double frequency, resulting in amplitude and phase oscillations of the  $\gamma^{(2)}$  signal as a function of the bond distance. As a consequence the two-body first shell terms are usually better accounted for by using the first terms of the corresponding MS expansions (up to  $\chi_4$  or  $\chi_6$ ).

In other cases the grouping procedure can proceed further and combine also  $\gamma^{(n)}$  signals with different  $n$  values involving sub-clusters of the main  $n$ -atom configuration. These latter signals are referred to as  $\eta$  signals. A typical example is that of a structure with a first coordination shell and a second coordination shell. The second shell atom will be first shell of the first shell, therefore each second shell distance is also part of a triangle with an additional intermediate first shell atom. In most of the cases the angle between the two bonds of this additional first shell atom is greater than 90 degrees and therefore the second shell  $\gamma^{(2)}$  and the three-body  $\gamma^{(3)}$  signals have very similar frequencies. The corresponding  $\eta$  signal is defined as the sum of those  $\gamma^{(2)}$  and  $\gamma^{(3)}$  signals and in practice is the effective additional signal of the second shell atom at all orders in the MS series.

The MS  $\chi$ ,  $\gamma$ , and  $\eta$  signals calculated by `gnxas`, as well as their low-order Taylor expansion coefficients, are used by the successive `fitheo` program to perform accurate MS data-analysis of EXAFS signals [1–3].

### 6.3. Program execution

#### 6.3.1. Input files

The program `gnxas` requires specific external input files produced by the program `phagen` in particular:

- A file containing the phase-shifts for the various atom (labeled 1,2,...) and angular momenta in a suitable energy mesh.
- A file containing the atomic cross-section  $\sigma_0$  for the various energies and both  $L_0+1$ ,  $L_0-1$  channels in the case of  $L_0 \neq 0$ .

All other information can be supplied manually into the input command file (see Appendix A3 for usage under different operating systems). It can be schematically subdivided into a head component, a central component, and a tail component. Header and tail components contain general information that specify calculation and output options. The central input component contains the geometrical information on the signals to be calculated, it can be input manually for simple cases or generated in nearly final form by the program `gnpeak`.

#### 6.3.2. Head component of input command file

This part is composed of 7 input cards as in the following example (the text after symbol ! are comments):

```

phase_shift.dat
sigma_0.dat
NIC          ! first three characters for output files
1            ! COORDINATES TYPE (1: Curvilinear)
3,8         ! mode + Xn limit
8.          ! RCUT
1           ! calculates derivatives? (0: No, 1: Yes, 2: second)

```

The detailed card description is as follows:

CARD 1: (`phase_shift.dat`) name of the phase shift external input file previously generated by `phagen`

CARD 2: (`sigma_0.dat`) name of the `sigma_0` external input file previously generated by `phagen`

- CARD 3: (ABC) A3 string used for the automatic generation of the names of the output files.
- CARD 4: (JC) Coordinates type: notice that the old option (JC=0: Cartesian) is now obsolete, use always JC=1, 3-body and 4-body peaks geometries should be specified using bond lengths and angles.
- CARD 5: (IW,NCUT) Mode calculation (IW) and maximum order of 'n' (NCUT) in the  $\chi_n$  calculation.  
 IW=1 Only total calculations (uses continued fractions).  
 IW=2 Only  $\chi_n$  signals.  
 IW=3 Both total signals and  $\chi_n$  signals are calculated.  
 IW=4 Same as 3 but optimized in the  $g_2$  case.
- CARD 6: (RCUT) Distance cutoff in the  $\chi_n$  path length.
- CARD 7: (JDER) Specify whether derivatives have to be calculated:  
 JDER=0 no derivative calculation, can be used to calculate the signals and produce appropriate plots;  
 JDER=1 derivatives are calculated and stored in the \*.DER files useful to perform a structural fit. For  $g_2$  second derivatives of amplitude are also calculated;  
 JDER=2 also second derivatives of phase in  $g_2$ .

### 6.3.3. Central part of input command file (basic usage)

This part contains the geometrical information for the 2, 3, and 4 body configurations to be calculated. For each of the three cases a specific card will specify the number of subsequent configurations included in the input file and the first and the last for which the calculation will have to be actually executed. In this way the same input file can be used with small modifications to calculate different signals still maintaining a independent nomenclature for the output files. An example of input is given by the following 5 cards:

```
2,1,2 ! number of g2 peaks, first and last peaks to be calculated
2.48678 12.00 1 3 1 2
3.51684 6.00 1 2 1 14
0,1,0 ! no g3 peaks
0,1,0 ! no g4 peaks
```

These cards refer to the input calculation for the first two pair contributions in a crystalline f.c.c. structure (example is Ni) The actual card description is as follows:

- CARD 8: (NG2,IIG2,IFG2 NG2) Number of listed peaks of the  $g_2$  distribution function (and of card 7 to follow); the calculations are performed from peak IIG2 to peak IFG2. Usually IIG2=1, IFG2=NG2.
- CARD 9: (R,A2,IP1,IP2,N1,N2) Defines the geometry of a two-body peak:  
 R Distance in Å;  
 A2 Coordination Number;  
 IP1,IP2 Phase shifts to be used for central and backscattering atoms relative to the phase shift file;  
 N1,N2 label for the central and backscattering atom are used only in plot files;

The successive two cards 0,1,0 in the previous example are needed to exclude any calculation on  $g_3$  and  $g_4$  peaks. These cards will be referred to as cards 10 and 12, respectively. The 0 character eliminates the need for input cards 11 and 13 containing geometrical information similar to card 9.

#### 6.3.4. Tail component of input command file

This component specifies additional calculation and output parameters. Other 6 cards are required as in the following example:

```

0           ! L=0    K-edge
.001,6,.00001,60 ! toll_lmax,mmax, Eps, N energy p.ts
3           ! IOUT
1.          ! k-weighting the plots
0.5,16.,.05  ! k_min, k_max, and dk of the interpolated signal
0.2         ! corresponding energy of the theoretical zero

```

CARD 14: L0 Angular momentum of the core-hole L0=0, K, L1, ...

CARD 15: (Toll\_lmax, Mmax, Eps, Nep) These are tolerance parameters which are set to optimize the calculation. Their explanation is given in the following paragraphs.

**TOLL\_lmax** It is a tolerance for the maximum angular momentum to be used at each energy. Starting from the maximum available angular momentum towards lower  $l$  values, the atom-energy dependent  $l_{\max}$  is fixed when:  $(2l+1) \cdot |t_l| > Toll_{l_{\max}} \cdot MM$ .  $MM$  is the maximum modulus of  $(2l+1) \cdot |t_l|$  among all atoms at the given energy. A typical value of the parameter is 0.001, values smaller than 0.01 may result in a coarse approximation of the spectrum which assumes a step-like shape at the  $l_{\max}$  change-points. Values of 0.0001 or smaller can be used if required.

**Mmax** It is the maximum  $M$  value used, independently of  $L_{\max}$ . Dropping large angular quantum numbers  $M$  yields an approximated treatment of the scattering when the angle is not zero or pi, however the contributions due to large  $M$ 's drop about one order of magnitude each time  $M$  is increased by one. A typical value for this parameter is 4, 5, or 6.

**EPS** It is the convergence parameter for the recursion algorithm. When the modulus of the difference between two successive continued fraction approximants is less than Eps the recursion is stopped. A typical value is in the range: .00001,.00000001.

**Nep** If smaller than the actual number of energy points in the mesh, the program will calculate only Nep energy points. Can be useful to test the calculation with a few points only.

CARD 16: (IOUT) General output flag:

```

IOUT < 3 No log_file written on the standard output.
      = 3 The detailed log_file of the continued fraction inversion is written
          on the standard output.

```

CARD 17: WH weight of  $k^{\text{WH}}$  in the signal plots,

CARD 18: (Kmin, Kmax, Dk) of the interpolated signal. These data affects only the plots, the actual calculation is performed for all energy points corresponding to the phase-shift mesh, possibly limited by the previous parameter Nep.

CARD 19: De energy shift of the theoretical zero for the  $k$ -scale (used only in the plot files). Notice: if the first point of the energy mesh has a negative energy this

shift should be sufficient to make it positive, otherwise a square root of negative value error may appear.

### 6.3.5. Central part of input command file (advanced usage)

An example of the central part of the input for several 2, 3, and 4 body peaks related to the previous f.c.c. Ni structure case is given below:

```

7,1,7 ! number of g2 peaks, first and last peaks to be calculated
2.48678 12.00 1 3 1 2
3.51684 6.00 1 2 1 14
4.30723 24.00 1 2 1 17
4.97356 12.00 1 2 1 16
5.56061 24.00 1 2 1 56
6.09135 8.00 1 2 1 69
6.57940 48.00 1 2 1 65
4,1,4 ! number of g3 peaks, first and last peaks to be calculated
2.48678 2.48678 60.00000 8.00 1 1 3.0 1 2 2 1 2 3
2.48678 2.48678 90.00000 12.00 2 1 1.0 1 2 2 1 2 5
                2 2.0 2 1 2 2 1 14
                6.000 ! two-body degeneracy II shell
2.48678 2.48678 120.00000 24.00 2 1 1.0 1 2 2 1 2 6
                2 2.0 2 1 2 2 1 17
                24.000 ! two-body degeneracy III shell
2.48678 2.48678 180.00000 6.00 2 1 1.0 1 2 2 1 2 11
                2 2.0 2 1 2 2 1 16
                12.000 ! two-body degeneracy IV shell
5,1,1 ! number of g4 peaks, first and last peaks to be calculated
'S' 2.48678 2.48678 60.000 2.48678 60.000 60.000 2.00 1
    1 4.00 1 2 2 2 1 2 3 4
'S' 2.48678 2.48678 60.000 2.48678 90.000 60.000 12.00 2
    1 2.00 1 2 2 2 1 2 3 7
    2 2.00 2 1 2 2 2 1 3 14
'S' 2.48678 2.48678 60.000 2.48678 120.000 60.000 12.00 2
    1 2.00 1 2 2 2 1 2 3 8
    2 2.00 2 1 2 2 2 1 3 17
'C' 2.48678 2.48678 90.000 2.48678 90.000 0.000 3.00 1
    1 4.00 1 2 2 2 1 2 15 5
'S' 2.48678 2.48678 60.000 2.48678 90.000 120.000 48.00 4
    1 1.00 1 2 2 2 1 2 3 5
    3 1.00 2 2 1 2 2 3 1 18
    2 1.00 2 1 2 2 2 1 3 15
    4 1.00 2 2 2 1 2 14 17 1

```

This input contains one input card 8 and seven input cards 9 previously described plus other cards described below in details.

CARD 10: (NG3, IIG3, IFG3 NG3) Number of peaks of the  $g_3$  distribution function. The calculations are performed from peak IIG3 to peak IFG3. Usually IIG3=1, IFG3=NG3.

CARD 11a: (R1, R2, TH, A3, NPo, PhP, Dgp, IP1, IP2, IP3, N1, N2, N3)

CARD 11b: (PhP, Dgp, IP1, IP2, IP3, N1, N2, N3)

CARD 11c: (Deg2) Defines the geometry of a three-body peak:

R1 distance between atom 1 and 2 in Å;  
R2 distance between atom 1 and 3 in Å;  
TH angle  $\widehat{213}$  in degrees;  
A3 Number of triangles/atom in the structure;



Npo Number of different photoabsorber positions;  
 PhP photoabsorber positions (1,2 or 3) in the triangle;  
 Dgp additional degeneracy of the position;  
 IP1, IP2, IP3 phase shifts to be used for atoms in positions 1, 2 and 3,  
 respectively, relative to the phase shift table contained in the  
 phase shift file;  
 N1, N2, N3 labels for the atoms in positions 1, 2 and 3, respectively. They  
 are used only in plot files;  
 Deg2 two-body degeneracy to calculate signal 2–3.

NOTE: *there is always 1 card of type 11a and Npo-1 cards of type 11b. Card 11c is required only if either Php=2 or Php=3 exist for the peak under consideration.*

CARD 12: (NG4, IIG4, IFG4 NG4) Number of peaks of the  $g_4$  distribution function. The calculations are performed from peak IIG4 to peak IFG4. Usually IIG4=1, IFG4=NG4.

CARD 13a: (TF, R1, R2, T1, R3, T2, T3, A4, Npo) Defines the geometry of a four-body peak.

→ TF: topological flag.

▷ If TF= 'S' star configuration, then:

R1 distance between atom 1 and 2 in Å

R2 distance between atom 1 and 3 in Å

T1 angle  $\widehat{213}$  in degrees

R3 distance between atom 1 and 4 in Å

T2 angle  $\widehat{214}$  in degrees

T3 angle  $\widehat{314}$  in degrees.

▷ If TF= 'C' chain configuration, then:

R1 distance between atom 1 and 2 in Å

R2 distance between atom 2 and 3 in Å

T1 angle  $\widehat{123}$  in degrees

R3 distance between atom 3 and 4 in Å

T2 angle  $\widehat{234}$  in degrees

T3 dihedral angle  $\widehat{1234}$  in degrees

A4 Number of quadruplets/atom in the structure

Npo Number of different photoabsorber positions. NOTE: *there is 1 card of type 13a for each peak.*

CARD 13b: (PhP, Dgp, IP1, IP2, IP3, IP4, N1, N2, N3, N4)

PhP photoabsorber positions (1, 2, 3 or 4) in the four-atom structure

Dgp additional degeneracy of the position

IP1, IP2, IP3, IP4 phase shifts to be used for atoms in positions 1, 2, 3 and 4,  
 respectively, relative to the phase shift table contained in the phase  
 shift file

N1, N2, N3, N4 labels for the atoms in positions 1, 2, 3 and 4, respectively. Are  
 used only in plot files.

NOTE: *there are Npo cards of type 13b for each peak.*

All previous cards contain geometrical and topological information that is rather difficult to determine manually. The input can be generated automatically using the `crymol` and `gnpeak` programs in sequence. This input part is similar to the `.gnx` output file of `gnpeak`. The users are anyway encouraged to check the content of the `.gnx` file prior to inclusion into the `gnxas` input, and in particular to limit the number of calculations. In any case the input lines corresponding to cards 8, 10, and 12 have to be completed.

#### 6.4. Description of the output files

Two types of output files are generated by `gnxas`: the binary `*.DER` files containing amplitude phases and derivatives of the various signals useful for the successive refinement of the XAFS signal (performed by the program `fitheo`) and the ASCII `*.PLO` files containing simple plots of the computed signals only useful for checking. The output file names are generated in an automatic way allowing for a quick and unambiguous identification. The names are composed of the A3 string ABC given in the input, that can be changed for different calculations, and are described in details below.

##### 6.4.1. Total plot files

These are formatted files suitable for producing a graphical output for workstations, terminals and postscript printers using the `TOPdrawer` program.

###### ABCGN.PLO

This file contains the successive contributions for the total signal associated with each peak of the various distribution functions in the input. In other words it is the plot of the total  $n$ -body  $\gamma$ -signals related to each input peak. If the input peak contains only one possible position of the photoabsorber then the total  $\gamma$ -signal coincides with the  $\gamma$ -signal referred to that position.

###### ABCG2n.PLO

###### ABCG3n.PLO

These files contain the various  $\chi_n$  signals associated with the  $n^{\text{th}}$  peak of the  $g_2$  or  $g_3$ . The extension 'n' is missing if only one peak is calculated. It is a I1 string if the number of peaks is smaller than 9 and is a I2 string if the peak is between 10 and 99. In the case of a  $g_3$  peak the file contains the various  $\chi_n$  signals related to the various different positions of the photoabsorber as well as the two-body  $\chi_n$  signals associated with the long bond of the triangle.

###### ABCG4n.PLO

This file contain the gamma 4 signals and possibly  $\gamma^{(3)}$  and  $\gamma^{(2)}$  signals associated with subunits. The standard sequence of the signals is P1G, P1S, P1F, P1L, P2G, P2S, P3G, P3S, P4G, P4S, P4F, P4L, TGA, TET. These A3 strings refer to the signal in the corresponding `*.DER` files, (see below). The extension 'n' is missing if only one peak is calculated. It is a I1 string if the number of peaks is smaller than 9 and is a I2 string if the peak is between 10 and 99.

##### 6.4.2. \*.DER files

These files are sequential access 16 records binary files which contain the basic information on the various  $\chi_n$  or  $\gamma$  signals. There are a few comment lines on top of each file followed by a line for each energy point. The variables stored for  $g_2$  cases are *Energy*,

Wavevector, Amplitude, Amplitude First and Second Derivative, Phase, and Phase First and Second Derivatives. In the case of  $g_3$  or  $g_4$  they are Energy, Wavevector, Amplitude, Amplitude Gradient, Phase, and Phase Gradient. These can be used to generate a signal at a slightly displaced position and to calculate configurational averages for narrow distributions of distances.

In the following file names the extension 'n' is missing if only one peak is calculated for the given distribution function. It is a I1 string if the number of peaks is smaller than 9 and is a I2 string if the number of peaks is between 10 and 99. The extension 'm' refers to the position of the photoabsorber, it can be 1, 2 or 3 for  $g_3$  peaks and 1, 2, 3 or 4 for  $g_4$  peaks.

→  $n^{\text{th}}$  peak of the  $g_2$  ( $n$  refers to the input sequence).

ABCG2n.DER is the total two-body signal ( $\gamma^{(2)}$ ) calculated by means of the continued fraction algorithm.

ABCG2nX2.DER is the corresponding single scattering  $\chi_2$  signal.

ABCG2nX4.DER is the corresponding multiple scattering  $\chi_4$  signal.

ABCG2nX6.DER is the corresponding multiple scattering  $\chi_6$  signal.

→  $n^{\text{th}}$  peak of the  $g_3$  ( $n$  refers to the input sequence).

ABCG3nPmG.DER is the total three-body signal  $\gamma^{(3)}$  calculated by means of the continued fraction algorithm with the photoabsorber in position 'm'.

ABCG3nPmL.DER exists either for m=2 or for m=3 if any of them is calculated and is the total two-body signal associated with the long bond of the triangular configuration, *i.e.* the bond opposite to the angle  $\theta$ .

ABCG3nPmE.DER is the sum of the three-body and the two-body signals and accounts for an effective signal of the distant (second) shell.

ABCG3nPmX301.DER Is the double scattering  $\chi_3$  signal with path  $0ij0$ .

ABCG3nPmX401.DER is the triple scattering  $\chi_4$  signal of path  $0i0j0$ .

ABCG3nPmX402.DER is the triple scattering  $\chi_4$  signal of path  $0ijj0$ .

ABCG3nPmX403.DER is the triple scattering  $\chi_4$  signal of path  $0jij0$ .

Other \*.DER files contain  $\chi_5$  and  $\chi_6$  signals.

ABCG3nTGA.DER is the signal  $\eta$  generated by summing the  $\gamma^{(3)}$  signals corresponding to the various positions of the photoabsorber. It is the one plotted in the ABCG3n.PLO file.

ABCG3nTET.DER is the signal generated by summing the  $\gamma^{(3)}$  signals corresponding to the various positions of the photoabsorber and the  $\gamma^{(2)}$  signal of the distant long bond. It can be helpful in collinear geometries with three atoms of the same type in order to use an effective peak signal instead of many separate ones.

→  $n^{\text{th}}$  peak of the  $g_4$  ( $n$  refers to the input sequence).

ABCG4nPmG.DER is the total four-body signal  $\gamma^{(4)}$  calculated by means of the continued fraction algorithm with the photoabsorber in position "m".

ABCG4nPmL.DER exists only in chain configuration if  $\text{Php}=1$  or 4. It is the total two-body signal associated with the long bond 1-4 of the four-atom configuration.

- ABCG4nPmS.DER exists only in chain configurations. If Php=1 or Php=2 it is the  $\gamma^{(3)}$  signal associated with atoms 1, 2, and 4. If Php=3 or Php=4 the atoms are 1, 3, and 4.
- ABCG4nPmF.DER exists only in chain configuration if Php=1, or 4. It is the  $\gamma^{(3)}$  signal associated with the atoms 1,3 and 4 or 1, 2, and 4, respectively.
- ABCG4nPmE.DER is the sum of the previous four-body, three-body and the two-body signals and accounts for an effective signal of the distant shell.

WARNING: In the case of four-body configurations the degeneracy of two-body and three-body sub-units is not accounted for independently. In the case of a highly connected network this degeneracy can be a fraction of the four-body degeneracy. The possibility of this occurrence should be checked manually, for example from the \*.gnp file generated by gnpeak and fractional degeneracy used correspondingly in the subsequent fittheo run.

## 7. Program fittheo

### 7.1. Introduction

The fittheo program is an advanced fitting routine developed to extract quantitative structural information by using the X-ray Absorption Spectroscopy (XAS). The first version of this program has been developed and used since 1990 in order to study the local structure of several systems (see references [1–3, 12] and GNXAS publication list in Appendix A2).

The program fittheo includes several innovative features for EXAFS structural refinements. The program is able to analyze raw x-ray absorption spectra without the need for any pre-treatment of the data (EXAFS extraction, Fourier filtering). Several spectra containing many absorption edges can be treated simultaneously, increasing the accuracy of the structural refinement. The atomic background modeling allows for the inclusion of specific features to account for the opening of many-electron absorption channels. Finally the program allows for a proper inclusion of “ab-initio” multiple scattering (MS) contributions in the EXAFS data-analysis. These MS contributions can be calculated by using the most advanced algorithms presently available within the one-electron and muffin-tin schemes (see gnexas documentation). The successful application of the MS theory to the XAS analysis has opened the possibility to determine quantitatively the parameters associated with short-range higher-order distribution functions as well as to improve the determination of the pair correlation function parameters (bond distances and variances).

The program uses (single and) multiple scattering contributions  $\chi_n(k)$  or  $\gamma^{(n)}(k)$  previously calculated for fixed atoms reference structures, using the appropriate ab-initio scheme (see gnexas documentation). Both MS expansions, meaningful only in the energy region of convergence, and total irreducible  $n$ -body multiple scattering signals, calculated by a continued fraction expansion method, can be used. Possible changes of the structural parameters from the reference geometries (such as distances, angles and coordination numbers) and the effect of the configurational (and thermal) average (over a Gaussian or non-Gaussian parameter distribution) result in specific parameter dependent amplitude and phase corrections that are applied to the input signals to generate a total model structural signal  $\chi_m(k)$ .

The program computes a model absorption signal  $\alpha_m(E)$  by adding to the structural oscillation  $\chi_m(k)$  a suitable background model. This signal is compared and refined directly

on the experimental absorption spectrum  $\alpha(E)$ . The non-linear multi-parametric fit is performed by using the MINUIT subroutine of the CERN library minimizing the summed squared residual between the model  $\alpha_m(E)$  and the experimental spectra  $\alpha(E)$  (refer to the MINUIT reference Manual [13]).

The basic assumption underlying the `fittheo` fitting scheme is that it is possible to account for the structural signal by adding a finite number of contributions associated with well defined peaks of the pair ( $g_2(r)$ ), three-body ( $g_3$ ), and four-body ( $g_4$ ) distribution-functions. This assumption is clearly valid for molecules and for crystalline solids where it is possible to define precise structural parameters for each peak and corresponding signal contributions. In the case of highly disordered systems, such as liquids or glasses, a peak fitting approach is questionable and alternative methods for their investigations have been developed, as described in the `fittheo`, `grfit`, `grxas`, `grrec` program documentation (see also GNXAS publication list, Appendix A2).

More details on the theory underlying this data-analysis method can be found in two detailed papers appeared consecutively in 1995 [1, 2]. The method was originally presented in 1991 [3]. Details on statistical treatment of the data and on multiple-edge data-analysis can be also found in references [12] and [14], respectively.

## 7.2. General description

### 7.2.1. Pre-analysis scheme

X-ray absorption spectra in the form of ASCII files containing absorption  $\alpha(E_i)$  as a function of the photon energy  $E_i$ , as they are normally collected at synchrotron radiation facilities, are suitable input data files for the program `fittheo` and do not require any pre-analysis or EXAFS extraction. For the optimal performance of the program the experimental spectra should be characterized by:

- a linear and clean pre-edge of at least 10 points;
- a well defined edge in such a way that the derivative of the absorption has a well defined maximum;
- a sufficiently fine energy sampling in the EXAFS region (*i.e.*  $\Delta k=0.02 \text{ \AA}^{-1}$ );
- the absence of pathological background artifacts.

In the case of a multi-edge refinements the files related to different atomic edges have to be supplied in order of increasing energy. A pre-analysis of each edge contained in the data files under consideration is performed at the beginning of the `fittheo` run (first call to the subroutine FCN in MINUIT, see ref. [13]). Such a preliminary analysis consist in the following actions:

- linear fit of the pre-edge region;
- calculation of the derivative of the absorption signal;
- determination of the experimental edge position, defined as the maximum of the first derivative of the spectrum;
- determination of the maximum of the absorption, called conventionally *white-line*;
- possibility of a reduction of the number of fitting points;
- evaluation of the statistical noise of the experimental spectrum;

The pre-edge linear fit does not represent the post-edge background in the EXAFS region. It is used only for the Jump evaluation of the absorption discontinuity at the edge (Jump, see below) and therefore it is extrapolated only at the edge energy value. The energy

limits of the pre-edge linear fit must exclude spikes and the beginning of the absorption edge. The fit is of the form  $a + E \times b$  where  $b$  represents the pre-edge slope in the edge region. The pre-edge linear fit is performed for all of the (NXAS) absorption edges under consideration.

The two-point derivative of the spectrum  $\alpha'(E'_i) = (\alpha(E_{i+1}) - \alpha(E_i)) / (E_{i+1} - E_i)$  ( $i$  runs over the energy point) is calculated from the first energy point up to a selected energy EMAX. The  $i$  point of the derivative is associated with the energy  $E'(i) = 0.5 \times (E(i) + E(i+1))$ .

The pre-edge slope is subtracted from the data. In this way the derivative spectrum is zero before the edge. The absolute maximum of the derivative in the calculated interval is taken as a first approximation of the edge energy  $EE$ .

The accuracy of the determination of the edge energy is controlled by a tolerance parameter TOLL (defined as a real number from 0 to 1). All the energy points, around the approximate edge energy  $EE$ , for which the derivative is greater of tolerance TOLL times the maximum derivative, are taken in account for a parabolic fit of the maximum of the derivative. If the number of points is less than 3 the nearest 3 points are used. In this way an accurate value for the experimental edge energy is found. If the spectrum is noisy with a coarse energy spacing a tolerance TOLL=0.9–0.99 is recommended, in such a way the edge is placed near the derivative maximum. If a smaller tolerance is used, due to the noise, many meaningless points are added which may move the maximum of the parabolic fit away from the edge region. Conversely if the spacing is fine, provided the derivative has a parabolic shape, a tolerance of 0.4–0.6 can be used which allow for an accurate determination of the maximum position, taken as edge energy  $EE$ . If needed,  $EE$  can be fixed to any specified energy value and no derivative is calculated.

With an additional parabolic fit the program determines the maximum of the absorption spectrum (minus the pre-edge linear fit) and its energy position: WL and EWL, respectively. This allows an immediate check of this first stage of analysis: for example in typical cases EWL is expected only a few eV above EE and WL is expected to be in the range 1 to 2.0–3.0 in Jump units.

After the execution of the pre-analysis described above there is the possibility of a reduction of the effective number of fitting points. This is useful to save CPU computer time during the first runs of `fittheo`. The reduction of the points is done interpolating the signals over a different mesh with constant step in the wave-vector space. This option can not be selected if a multiple-edge fitting is attempted. The usefulness of reducing the number of points can be appreciated if the total number of fitting points becomes more, let's say, than 500: in this case usually one can save one half of CPU time simply accepting a mesh spaced 0.05 in  $k$ -space. The final runs of the fitting procedure can be done in any case using the original data.

An estimate of the noise of the data can be performed over a selected number of points of the experimental spectrum. If a multiple-edge fitting is attempted, noise can be evaluated separately for each absorption spectrum under consideration. The evaluation of noise is done by calculating the sum of the square differences between data and best-fit polynomial functions. Results are plotted using a logarithmic scale. The average standard deviation(s) of the measured absorption coefficient(s) as well as the expected value for the residual function to be minimized, are estimated. The residual function is a  $\chi^2$ -like function, and its expected value depends on the variance of the experimental data. Estimate of the noise

of the measured data is particularly useful for evaluating appropriate confidence intervals for the structural parameters refined using the fitting procedure (see Section 7.2.3). A final useful calculation performed by the `errstat` subroutine in `fittheo` is the identification of suspected glitch points, detected as points above the average noise level, and visualized in a table.

A preliminary pre and post-edge background extraction without considering the model structural signal is performed prior to the first minimization by an appropriate subroutine. There is no need for introducing background coefficients as MINUIT parameters, the polynomial coefficients (splines) are automatically optimized at each MINUIT call. It is possible to FIX and use spline coefficients as MINUIT parameters by choosing an appropriate option in the spline menu (see the next subsection).

### 7.2.2. Background model

As anticipated, the program compares the experimental absorption spectrum  $\alpha(E_i)$  with a model one  $\alpha_m(E_i)$ , where  $i$  runs over the energy points of the experimental data. The model is composed of a background part and the structural signal  $\chi_m(k)$ . The wave vector values  $k$  are of course related to the energy values  $E_i$ . The background is given by the atomic absorption  $\alpha_0(E_i)$ , the post-edge background  $\alpha_{bkg}(E_i)$  and sometimes by the contribution of additional edges  $\alpha_{exc}(E_i)$  in the EXAFS region. The atomic absorption is used only to normalize the structural signal  $\chi_m(k)$  and is eventually included in the post-edge background function  $\alpha_{bkg}(E_i)$ . The additional edges contained in  $\alpha_{exc}(E_i)$  are usually associated with many-electron excitations, neglected in present one-electron calculations, whose amplitudes can be of the order of 1% of the main one-electron channel. Therefore the model signal is defined as:

$$\alpha_m(E_i) = \alpha_{bkg}(E_i) + \alpha_{exc}(E_i) + \chi_m(E_i)\alpha_0(E_i)$$

that is the quantity to be compared with the experiment. In case of multiple-edge fitting, the model signal is:

$$\alpha_m(E_i) = \alpha_{bkg}(E_i) + \alpha_{exc}(E_i) + \sum_{j=1}^{NXAS} \chi_m^j(E_i)\alpha_0^j(E_i)$$

where the sum  $j = 1, NXAS$  runs over the specified number `NXAS` of atomic edges.

**7.2.2.1. Post-edge background.** The background is the atomic contribution of the particular edge under study plus the (continuation of the) pre-edge background (post-edge background of other edges) which is not accounted for by the linear fit. Only the atomic contribution is used to normalize the EXAFS  $\chi(k)$  oscillation. The background is composed of a given number of polynomials (for defined energy intervals) connected with continuity and derivability (but with the second derivative discontinuous). Starting from a specified MINUIT call a post-edge background extraction is automatically done for each change in the parameter space performing a least-squares fit on the appropriate function. The post-edge background refinement can be also stopped by the user during the minimization. By default, the post-edge background is always optimized. In any case, there is the possibility to assign the background parameters as MINUIT fitting parameters which can be optimized in the minimization together with the structural ones. In this way the convergence to the minimum is generally slower but in some cases it is possible to obtain a better minimum. If the latter option is used, the MINUIT parameters are allocated in this way: if the first

polynomial has degree  $n$ , the MINUIT parameters  $X(1), X(2) \dots X(n+1)$  are used, in the same way the second of degree  $m$  uses  $X(n+2) \dots X(n+m)$ . Indeed the first has  $(n+1)$  degrees of freedom, the second  $(m-1)$  because of the two constraints at the connecting point, and so on.

The post-edge background is therefore defined by a function  $\alpha_{exc}(E)$ . This smooth function, which is fit in the above mentioned energy interval, is extrapolated at the edge energy providing an estimate of the edge Jump. The edge Jump is defined as the absorption discontinuity between the linear pre-edge and the smooth post-edge background extrapolated at the previously determined edge energy  $EE$ . Since the edge energy is taken as origin in the successive analysis the first background parameter is directly the Jump of the spectrum. It is advisable to work with a limited number of spline parameters. A number of 4 or 5 over 200 energy points can be a reasonable choice. This works unless the background shape actually requires a larger number of parameters. It is important to remark that the smoothness of the post-edge background  $\alpha_{bkg}(E)$  has to be verified using the graphical output of the program (see Section 7.3.2, PA'ABC'.DAT output file), in order to avoid major mistakes associated with possible unphysical oscillations of the  $\alpha_{bkg}(E)$  function.

As anticipated, the spline parameters could also be kept fixed in the minimization procedure if the background model were considered accurate. A further possibility of the post-edge background algorithm consists in an edge constraint in the spline parameters. The linearized hydrogenic model for the atomic background absorption of a particular edge predicts the following behavior right after the threshold:

$$\alpha_0(E_i) = J \times \left[ 1 - \frac{8}{3} \times \frac{(E_i - EE)}{EE} + O((E_i - EE)^2) \right]$$

This means that at the edge the first derivative of the background with respect to the energy is just  $\frac{8}{3} \frac{J}{EE}$ . This is a constraint for the ratio of the first two spline coefficients. The number of spline parameter can be therefore reduced by 1 calculating the second coefficient as  $\frac{8}{3} \frac{J}{EE}$  simply answering Y to the appropriate question. For multiple-edge fitting this choice can be selected for each edge under consideration, separately. If the pre-edge region is not reproduced correctly or if the particular edge background departs from this behaviour this constraint should not be applied.

*7.2.2.2. Atomic background, normalization of the structural signal.* The use of the model signal  $\chi_m(k)$  and the normalization of the  $\chi(k)$  spectrum is performed by using an appropriate model function for the atomic absorption. The formula that we use is the well-known hydrogen-like absorption function (see for example ref. [15]) and defines the energy dependence of the atomic cross-section  $\sigma_{nor}(E_i)$ ,  $\sigma_{nor}(E_i) \times J = \alpha_0(E_i)$ :

$$\sigma_{nor}(E_i) = \frac{1}{e^{-4}} e^{(-4 \times \arctan \frac{1}{\eta})} \times \frac{1}{\left(1 + \frac{1}{\eta^2}\right)^4} \times \frac{1}{1 - e^{-2\pi\eta}}$$

where  $\eta^2 = \frac{EE}{(E_i - EE)}$ . For  $\eta \rightarrow \infty$  ( $E_i \rightarrow EE$ ) the Taylor expansion of this formula gives the previously defined  $\alpha_0(i)/J$  near the edge. This normalization function is computed for all of the energy points of the edges under consideration.

The experimental structural signal is therefore defined as:

$$\chi(E_i) = \frac{[\alpha(E_i) - \alpha_{bkg}(E_i) - \alpha_{exc}(E_i)]}{\sigma_{nor}(E_i) \times J}$$



When more than one edge is considered ( $j = 1, \text{NXAS}$ ), we have:

$$\chi(E_i, j) = \frac{\left[ \alpha(E_i) - \alpha_{bkg}(E_i) - \alpha_{exc}(E_i) - \sum_{l \neq j} J_l \sigma_{nor}(E_i, l) \chi(E_i, l) \right]}{\sigma_{nor}(E_i, j) \times J_j}$$

defined only for energies greater than the edge energy  $EE(j)$  of the edge under consideration. The sum ( $\sum_{l \neq j}$ ) over the other structural signals obviously excludes the  $j$  edge under consideration and is not performed if the edge energy  $EE(l)$  of the  $l^{\text{th}}$  edge is greater than the edge energy  $EE(j)$ . Important effects due to the continuation of the structural signals of absorption edges previously opened can be found in  $L$ -edges monatomic studies of fifth-period elements [12] and in many multiatomic systems.

*7.2.2.3. Inclusion of many-electron excitations in the EXAFS region.* The presence of background features associated with the onset of double-electron excitation channels is now well established. These features have been identified in most of the low-noise  $K$ -edge and  $L$ -edge spectra. A noticeable example is given by the  $KL$  ( $[1s2p]$ ) edge (visible in the EXAFS region of the  $K$ -edge) of third period elements for which photomission studies have clearly revealed the corresponding shake-up satellites.  $KM$  edges ( $[1s3p]$  and  $[1s3d]$ ) were found to be important features in various  $Br$   $K$ -edge spectra.  $LN$  [ $2p(s)4f$ ] edges in  $L$ -edge spectra of sixth-period elements and of  $KM$  [ $1s3p$ ]/ $[1s3d]$  edges in  $K$ -edge spectra of fifth-period elements were also identified. A rather complete list of papers published by our group reporting evidences of double-electron excitation channels can be found in the GNXAS publication list (Appendix A2).

These double-electron excitation channels are, in general, responsible for weak background anomalies of the order (at most) of a few percent of the main edge jump. Their intensity may sometimes compete with that of the EXAFS structural signal. The most typical shape resembles a small smoothed edge step often accompanied by a background slope change and sometimes visible discrete threshold autoionizing resonances. Current theoretical approaches are unable to calculate accurate model shapes and intensities for these absorption channels. However, the energy onset of the channels can be predicted with relatively high accuracy by means of the energy differences between self consistent atomic calculations.

Within the presently available data-analysis scheme these features can be accounted for only at an empirical level. For this purpose several functional models, able to model additional steps and slope changes at selected energies in the background function, are included in the `fittheo` program. Several of these empirical models have been developed and successfully used to take into proper account the contribution of double-electron excitation channels into XAS data (see GNXAS publication list and references therein, Appendix A2). The main model functions currently available in the `fittheo` program include:

$$\begin{aligned} \alpha_{exc}(E) &= J \left[ \frac{\arctg\left(\frac{E'}{\Delta E}\right)}{\pi} + \frac{1}{2} \right]; & 'A' \\ \alpha_{exc}(E) &= J \left[ 1 - \frac{8}{3} \left( \frac{E'}{\Delta E} \right) \right] \left[ 1 - \left( \frac{\Delta E}{E' + \Delta E} \right)^2 \right], \quad E' \geq 0; & 'E' \\ \alpha_{exc}(E) &= \begin{cases} 0, & E' \leq 0; \\ \frac{p}{4\Delta E} (E' + \Delta E)^2, & -\Delta E \leq E' \leq \Delta E; \\ p \times E', & E' \geq \Delta E; \end{cases} & 'P' \end{aligned}$$

$$\alpha_{exc}(E) = J \frac{1}{1 + \left(\frac{E'}{\Delta E}\right)^2}; \quad \text{'L'}$$

$$\alpha_{exc}(E) = \begin{cases} 0, & E' \leq -2\delta; \\ \frac{2J}{\delta} \frac{(E'+2\delta)^2}{4\delta}, & -2\delta \leq E' \leq 0; \\ [J + E'J'] \left[1 - \left(\frac{\Delta E}{E'+\Delta E}\right)^2\right] + \frac{2J\delta}{\Delta E}, & E' \geq 0; \end{cases} \quad \text{'F'}$$

The first four functions depend on three parameters (Jump  $J$  or slope  $p$ , width  $\Delta E$ , and onset energy  $E_D$ , where  $E' = E - E_D$ ) and can be selected using the appropriate flag T\_ED (card 7.1) in the `fitheo` command file. T\_ED is A, E, P, and L for the arctangent-like, smoothed step, smoothed slope change, and Lorentzian functions, respectively. The last function is activated by the flag F and depends upon five parameters defining a more complex smoothed slope change with the final slope  $J'$  and two widths  $\delta$  (initial) and  $\Delta E$  (final) for the double-electron channel. Other empirical functions are currently supported by `fitheo`, some of their equations are published (see GNXAS publication list, Appendix A2) or can be provided upon request.

The A, E, P, and F functions are used in the cases of smooth onsets, and one can specify or refine the energy position, the onset interval and the jump in the absorption. The Lorentzian function can be used in the case of visible resonances. These empirical additional background functions can reproduce the asymptotic trend of a shake-up channels or the main lineshape of a resonance. Usually the onset region is not reproduced correctly and it should be excluded from the fit. In general, the inclusion of these contributions was demonstrated important since it allows for a larger background flexibility, able to model slope changes and discontinuities around the onset of multi-electron excitation edges, thus improving the agreement in the refinement of the structural signal (EXAFS).

### 7.2.3. Minimization procedure, statistical analysis

The function  $R$  which is minimized in the program is the sum over  $i = 1, N$  points of the squares of the differences between model and experimental signals in the specified energy interval times a weight function:

$$R = \sum_i [\alpha(E_i) - \alpha_m(E_i, \{\lambda\})]^2 \times W(E_i)$$

where the model signal  $\alpha_m(i)$  depends on the ensemble of structural parameters  $\{\lambda\}$ . The weighting function  $W(E_i)$  is calculated as:

$$W(E_j) = \frac{N}{N-p} \times \frac{k_j^w}{\sum_i [\alpha_N(E_i)^2] \times k_i^w}$$

where  $k_i (\text{\AA}^{-1}) \sim 0.512\sqrt{(E_i - EE)(\text{eV})}$  and  $p$  is the number of parameters used in the minimization.  $\alpha_N(E_i) = \alpha(E_i) - a - b \times E_i$  is the absorption spectrum obtained removing the pre-edge linear background (see Section 7.2.1).

Under normal conditions the weighting function should mimic the energy dependence of the inverse noise variance  $1/\sigma_i^2$ . This can be achieved by tuning the  $w$  parameter until a satisfactory agreement with the estimated noise is obtained, as reported in a suitable graphical output (see Section 7.3.2.3). When the structural model is able to reproduce completely the experimental signal (apart from the noise) the expected value for the residual function  $R$  represents a weighted average squared noise (usually in the range  $10^{-06}$ – $10^{-08}$ ).

In most of the real situations the actual minimum value of the residual will exceed these limiting values due to the presence of unexplained signal contributions. Moreover the approximations in the signals calculations are an additional source of discrepancy which generally increases the achievable minimum value. In common practice values of the  $w$  parameter in the range 1–4 are used. The larger values have the effect of giving a larger weight to higher energy data, and this choice can be adopted to reduce the weight of low energy data where the theoretical model may be less accurate.

Often, there are some points or particular regions of the experimental spectra affected by much larger noise, like spikes, glitches or double excitation edges. Several energy intervals can be easily set with weight  $W(E)=0$  and therefore excluded from the minimization.

When a multiple-edge fitting is attempted, minimization is similarly performed. The weighting factors  $W(E_i, j)$  for each experimental signal  $\alpha(E_i)$  in the energy region associated with the particular measurement ( $j=1, \text{NXAFIL}$ , total number of files containing raw absorption data) considered in the multiple-edge fit, can be different according to the different experimental conditions. The actual weight of each absorption edge spectrum is proportional to the inverse of the variance  $\sigma^2$  associated with each experimental spectrum.

The variance of the experimental data  $\sigma_{\text{expt}}^2$  can be automatically calculated but the use of different weighting factors is allowed. The use of a power-like functional form for the weight,  $k^w$ , can be applied to the various edges under consideration. The expected value of the residual function  $R$  is also calculated.

The above-mentioned residual function  $R$  follows the  $\chi_N^2$  distribution, provided that the weight  $W(E_i)$  is proportional to the inverse of the variance  $\sigma^2$  of the  $[\alpha(E_i) - \alpha_m(E_i)]$  random variable. Under these conditions, a full statistical evaluation of the results can be performed:

- 1) within a given choice for the structural model, the optimal values for the  $\{\lambda\}$  structural parameters are the set  $\{\lambda_{\text{min}}\}$  such that the residual  $R$  is at a minimum.
- 2) The statistical  $\chi^2$  test can be performed to check whether the actual value of  $R$  is only due to residual noise or it contains unexplained physical information.
- 3) A comparison between two models containing a different number of structural parameters can be performed on the basis of the  $F$ -test, applicable on  $\chi^2$  distributions. For instance, it can be verified if the reduction of  $R$  obtained using more parameters is statistically significant.
- 4) Statistical errors related to selected confidence intervals can be estimated for the structural parameters, looking at regions in the parameter space for which  $R(\{\lambda\}) < R_{\text{min}} + C$ , where  $C$  depends on the confidence level chosen and  $R_{\text{min}}$  corresponds to the expected value (when the residual contains only statistical noise). These regions, in the second-order approximation, are  $p$ -dimensional ellipsoids which provide also an insight onto the correlation among parameters.  $R(\{\lambda\})$ , for different realizations of the experimental noise, follows a  $\chi_p^2$  distribution where  $p$  is the number of parameters (see for example pages 684-694 in “Numerical Recipes in Fortran 77”, ref. [16]). The value for  $C$  related to a 95% confidence level can be then approximated by  $C \sim R_{\text{min}} \times [p + 2\sqrt{(2p)}]/N$ .

As anticipated, the non-linear minimization in parameters' space is performed by using the MINUIT subroutine of the CERN library (see ref. [13] for details). Several different

minimization routines can be used using different options of the MINUIT subroutine. Parameters can be optimized in selected ranges or kept fixed at various stages of the minimization. The ERRdef value should be of the same order with the expected residual  $E\{R\}$  for best results. Statistical errors on the structural parameters can be estimated defining the ERRdef value  $C$  according to the selected confidence level (see also the documentation about the Estimation of statistical errors, Section 8).

#### 7.2.4. Structural parameters

The fit performs the refinement of a series of parameters associated with the distribution of interatomic distances and angles around a selected atomic site. Due to the peculiar characteristics of the XAS signals, only short-range structural parameters can actually be refined. The relevant structural parameters are defined by appropriate MINUIT parameter cards.

A  $n$ -body decomposition of the structural signal, which has been shown to have good convergence properties, [1] is used to define the relevant structural parameters. In real vibrating or disordered systems, XAS spectra are sensitive to the short-range part of the first  $n$ -body distribution functions. The fitting procedure takes advantage of the precise relationship between MS signals and  $n$ -body distribution functions.

Let us consider the relative distribution of atom  $o$  and  $i$  which is a pair distribution function  $g_2(r)$ . This distribution is formed by a series of peaks which can be just one like in a diatomic molecule, several ones like in a larger molecule, cluster or crystal, and a few merging into a continuum like in amorphous solids or liquids. In the following, we assume that in any case the relevant features of the  $g_2(r)$  are isolated peaks which can be modelled using a Gaussian function or alternatively a skewed distribution with a given third cumulant. However, there are specific methods to treat the case of highly disordered systems where a broad first-neighbour merges into a long range tail (like in liquid metals).

Each of the short-range  $g_2(r)$  peaks used to define the short-range structure is associated with particular geometrical configurations and is specified by number of atoms ( $N$ ), average distance ( $R = \langle r \rangle$ ), variance ( $\sigma^2$ ) and third cumulant ( $C_3$ ). The  $g_2(r)$  peak contributes with signals due to single scattering paths of the type  $o-i-o$  ( $\chi_2$ ) or MS paths of the type  $o-i-o-i-o$  ( $\chi_4$ )...according to the MS expansion (see `gnxas` documentation, Section 6).

The total signal due to a specified  $g_2$  peak, defined as the irreducible two-body XAS signal  $\gamma_2(k)$ , can be calculated using the continued fraction expansion technique and used in the fitting procedure (if the series converges  $\gamma_2(k)$  is equal to the sum of MS terms  $\chi_n$ , with even  $n$ ).

In an analogous way the relative distribution of atoms  $o$ ,  $i$ , and  $j$ , which is the three-body distribution  $g_3$ , can be represented by the sum of a finite number of peaks modelled by appropriate mean values (three coordinates  $\{r\}=R_1, R_2$  and  $\theta$ ), covariance matrices and third cumulants. Well-defined  $g_3$  peaks are observed in crystals, molecules and covalent disordered systems. Each of these peaks contributes with the MS signals  $\chi_3$ :  $o-i-j-o$ ,  $o-j-i-o$ , and  $\chi_4$   $o-i-o-j-o$ ,  $o-j-o-i-o$ ,  $o-i-j-i-o$ ,  $o-j-i-j-o$  +  $O(\chi_5)$ .

In this case it is even more convenient to use directly the total signals  $\gamma_3(k)$  or  $\eta_3(k)$  (which takes account also of the two-body contribution related to the more distant atom) in order to avoid the introduction of too many contributions.

Similarly, the relative distribution of atoms  $o, i, j, k$  which is the four-body distribution  $g_4$ , can be represented by the sum of a finite number of peaks. Four-body configurations depend, in general, on six different coordinates. Four-body  $\gamma_4(k)$  contributions are generally weak, but strong signals can be associated with quasi-linear four-body configurations.

The general expression involving MS  $\gamma_n(k, \{r\})$  signals and  $n$ -body distribution functions  $g_n(\{r\})$  in the definition of the XAS  $\chi_m(k)$  structural signal is the following:

$$\chi_m(k) = \int dr 4\pi r^2 \rho g_2(r) \gamma_2(k, r) + \int dr_1 dr_2 d\theta 8\pi^2 r_1^2 r_2^2 \rho g_3(\{r\}) \gamma_3(k, \{r\}) + \dots$$

where the XAS short-range sensitivity is due to the strong amplitude decay of the  $\gamma_n(k, \{r\})$  increasing the distances (defined by the coordinates  $\{r\}$ ) associated with the  $n$ -body configuration. In practice, limiting ourselves to the Gaussian case, the parameters that have to be introduced for each peak of the  $g_2$  and  $g_3$ , are:

- $g_2$  : each peak is defined by
  - one distance  $R$  (Å),
  - one coordination number  $N$ ,
  - one variance  $\sigma^2$  (Å<sup>2</sup>),
- $g_3$  : each peak is defined (in angular coordinates) by two distances  $R_1, R_2$  and one angle  $\theta$  (deg) defining a triangle,
  - one factor defining the multiplicity of the peak,
  - one 6-parameter symmetric covariance matrix containing bond, angle variances, bond-bond and bond-angle correlations defined as:

$$\begin{pmatrix} \sigma_{R_1}^2 & \sigma_{R_1 R_2}^2 & \sigma_{R_1 \theta}^2 \\ - & \sigma_{R_2}^2 & \sigma_{R_2 \theta}^2 \\ - & - & \sigma_{\theta}^2 \end{pmatrix}$$

The fitting parameter used by `fittheo` are the following:

$$\begin{pmatrix} \sigma_{R_1}^2 & \rho_{R_1 R_2}^2 & \rho_{R_1 \theta}^2 \\ - & \sigma_{R_2}^2 & \rho_{R_2 \theta}^2 \\ - & - & \sigma_{\theta}^2 \end{pmatrix}$$

where

$$\rho_{ij} = \frac{\sigma_{ij}^2}{\sqrt{(\sigma_i^2 \times \sigma_j^2)}} \quad \text{with } i, j = R_1, R_2, \theta \quad \text{and } i \neq j$$

in this way  $\rho_{ij}$  is dimensionless and defined in the range  $(-1, 1)$ .

- $g_4$  : each peak is defined (in angular coordinates) by:
  - three distances  $R_1, R_2, R_3$ ,  $\theta$  ( $R_1 - R_2$ ),  $\phi$  ( $R_2 - R_3$ ) bond angles, and  $\psi$  (deg); those angles have been indicated as  $\theta_{123}$ ,  $\theta_{234}$ ,  $\phi$  (“chain” topology) and as  $\theta_{213}$ ,  $\theta_{314}$ ,  $\theta_{214}$  (“star” topology) in the `gnpeak` documentation.
  - one factor defining the multiplicity of the peak;

– one 21-parameter symmetric covariance matrix defined as

$$\begin{pmatrix} \sigma_{R_1}^2 & \sigma_{R_1 R_2}^2 & \sigma_{R_1 R_3}^2 & \sigma_{R_1 \theta}^2 & \sigma_{R_1 \phi}^2 & \sigma_{R_1 \psi}^2 \\ - & \sigma_{R_2}^2 & \sigma_{R_2 R_3}^2 & \sigma_{R_2 \theta}^2 & \sigma_{R_2 \phi}^2 & \sigma_{R_2 \psi}^2 \\ - & - & \sigma_{R_3}^2 & \sigma_{R_3 \theta}^2 & \sigma_{R_3 \phi}^2 & \sigma_{R_3 \psi}^2 \\ - & - & - & \sigma_{\theta}^2 & \sigma_{\theta \phi}^2 & \sigma_{\theta \psi}^2 \\ - & - & - & - & \sigma_{\phi}^2 & \sigma_{\phi \psi}^2 \\ - & - & - & - & - & \sigma_{\psi}^2 \end{pmatrix}$$

Also in this case the actual bond-bond and bond correlations  $\sigma_{ij}^2$  are replaced by dimensionless parameters

$$\rho_{ij} = \frac{\sigma_{ij}^2}{\sqrt{(\sigma_i^2 \times \sigma_j^2)}} \quad \text{with } i, j = R_1, R_2, R_3, \theta, \phi, \psi \quad \text{and } i \neq j$$

defined in the range  $(-1, 1)$ .

The number of structural parameters that has to be introduced is actually limited because  $g_2$  and  $g_3$  peaks are hierarchically interconnected both for what concerns the average quantities (distances) and for the variances ( $\sigma_i^2$ ). The program allows one to use the same structural parameters (for example a  $g_2$  first-shell distance and variance  $R_1, \sigma_{R_1}^2$ ) for all the  $g_3, g_4$  peaks defining that structure. The inclusion of higher-order peaks can improve the accuracy in the determination of structural parameters.

In most cases, the actual number of parameters used to define a  $g_3$  peak is reduced by symmetry. For example, a  $g_3$  peak related to isosceles triangles will depend on two average quantities (distance  $R$  and angle  $\theta$ ) and four covariance parameters only ( $\sigma_{R_1}^2, \sigma_{R_1 R_1}^2, \sigma_{R_1 \theta}^2, \sigma_{\theta}^2$ ).

For an equilateral configuration the number of parameters is reduced even more. By definition, there is only one average distance  $R$  (the average angle being 60 degree), and the covariance matrix contains only two different parameters: one bond variance ( $\sigma_{R_1}^2$ ), and one bond-bond correlation ( $\sigma_{R_1 R_1}^2$ ) defining completely the vibration of the triangle.

Specialized expressions have to be used in the extremal case of collinear three-body configurations. First-derivatives of phases and amplitudes of the MS signals with respect to the angle vanish and second-order expansion has to be used [17]. The three-body distribution vanishes at exactly 180 degrees and can be modelled as described in [2]. The distribution depends in general on six parameters: two bond distances  $R_1$  and  $R_2$ , two bond variances ( $\sigma_{R_1}^2, \sigma_{R_2}^2$ ), the bond-bond correlation ( $\sigma_{R_1 R_2}^2$ ), and the angle variance. When only one bond distance is defined the total number of parameters reduces to four.

Usually, the  $n$ -body expansion converges quite rapidly and the inclusion of four-body signals can be avoided. However, quite intense MS contributions can be associated with four-body quasi-collinear configurations present in the local structure. In most cases, it is not necessary to define the entire 21-parameter covariance matrix (and six average coordinates) because specialized expressions are also used for  $g_4$  peaks involving collinear or planar configurations. In particular, for “chain”-type four-body collinear configurations ( $\theta = \phi = 180^\circ$ ) we need to define only three average quantities  $R_1, R_2, R_3$  (the average dihedral angle  $\psi = 0$  can be assumed to be zero). Deviations from planarity can be neglected and the four-body distribution factorizes into a distribution of the  $R_1, R_2$ , and  $R_3$  distances and the distribution of  $\theta$  and  $\phi$  angles around 180 degrees. The original 21 parameters

covariance matrix reduces to only 8 independent parameters ( $\sigma_{R_1}^2$ ,  $\sigma_{R_2}^2$ ,  $\sigma_{R_3}^2$ ,  $\sigma_{R_1 R_2}^2$ ,  $\sigma_{R_1 R_3}^2$ ,  $\sigma_{R_2 R_3}^2$  and the two angle variances). When two or more bond lengths coincide the number of parameters is even smaller.

Other specialized expressions are used for  $g_4$  peaks where partial linearity ( $\theta$  or  $\phi$  equal to 180 degrees) or simple planarity is present. For a generic partial linear configuration, the average quantities are 4 ( $R_1$ ,  $R_2$ ,  $R_3$ , and one angle,  $\theta$  or  $\phi$ ) and the covariance matrix contains 11 independent parameters (five variances, plus three bond-bond and three bond-angle correlations). For a generic planar configuration ( $\psi = 0$ ), there are 5 average quantities ( $R_1$ ,  $R_2$ ,  $R_3$ ,  $\theta$ , and  $\phi$ ) and the covariance matrix contains 16 independent parameters (only the cross terms involving  $\phi$  are excluded).

In any case, the total number of parameters is always kept to a minimum taking advantage of the hierarchical relationship between  $g_4$ ,  $g_3$  and  $g_2$  peaks (the same parameters are used in the definition of various interconnected peaks).

In several cases, the bond-length distribution deviates from that described by a simple Gaussian model. This happens in crystalline solids at high temperatures and in practically all of the solid and liquid disordered systems. In many interesting cases, the harmonic model fails even at moderate temperature. In all these cases, a more flexible model for the distribution function is required. For the pair distribution function  $g_2$ , a model which has been proven to be useful is the  $\Gamma$ -like distribution reported in ref. [18] and [19], and briefly described in Section 9

The  $\Gamma$ -like function depends on the average distance  $R$ , variance  $\sigma^2$  and a dimensionless skewness  $\beta$ . The limit for  $\beta = 0$  is the Gaussian distribution. This function has been shown to reproduce  $g(r)$  peak shapes obtained using computer simulations in various systems.

The parameter  $\beta$  is directly related to the third cumulant

$$\beta = \frac{C_3}{\sigma^3}.$$

The other cumulants of the distribution are always defined given the functional form.

The cumulant expansion including the  $C_3$  and  $C_4$  terms can always be used for the  $g_2$ ,  $g_3$  and  $g_4$  peaks. The series is slowly convergent and can be useful only for small deviations from the Gaussian approximation. Moreover, the reconstruction of the  $g(r)$  shape is not always possible for a given set of cumulants. Large statistical errors in the values of the higher-order cumulants generally lead to unacceptable uncertainties in the  $g(r)$  shape. Therefore, the use of the cumulant expansion is usually *not recommended*.

Of course, the effective number of two-body, three-body and four-body equivalent configuration (multiplicity) has to be defined too. Several coordination numbers  $N$  can be introduced in the fitting procedure and used for different MS  $\chi_m$  ( $m \geq n$ ),  $\gamma^{(n)}(k)$  or  $\eta^{(n)}(k)$  signals associated with the  $g_n$  peaks.

Indeed, when the relative multiplicity is known, it is possible to use the same coordination number for different atomic configurations, by introducing suitable multiplicative factors accounting for the actual number of atomic configurations. It is also possible to introduce physical constraints. For example, the sum of the coordination number of two shells can be kept fixed ( $N_1 + N_2 = N$ ). Moreover, if a simultaneous multiple-edge fitting procedure is performed, some structural parameters can be relevant to one or more edges. The MS

signals related to a selected  $g_2$ ,  $g_3$ , or  $g_4$  peak are labelled according to the particular edge for which they were calculated.

More sophisticated constraints are possible, and an important one is related to the use of the present “peak-fitting” approach to XAS studies of disordered systems. In typical disordered systems like liquids, glasses and amorphous materials the above mentioned  $g_n$  peaks are not separated in the coordinates’ space. For the pair distribution function  $g_2(r)$  it has been shown that appropriate constraints for the peak parameters have to be used in order to satisfy the correct thermodynamic limit (see ref. [20]). The use of such constraints described in the `grxas` documentation has been proven to be necessary for reliable structural refinements of liquid systems (see also the GNXAS publication list, Appendix A2).

#### 7.2.5. Other parameters

The present theory of X-ray absorption contains some other physical quantities in the calculation of the cross-section that have no direct structural meaning. The nature of these quantities is mainly related to the description of the interaction of the photoelectron with deep and valence electrons. The introduction of phenomenological quantities describing the inelastic processes, the core-hole lifetime and many-electron effects is in any case necessary to obtain a good fit of the experimental spectra. The finite energy resolution of the experimental spectra has to be also taken into account in the calculation of the model signal. The energy resolution is usually known with reasonable accuracy and can be fixed to the known value in the fitting procedure. The difference between the theoretical and the experimental energy scale  $E_{0T}$  has to be also refined. In standard situations, only  $S_0^2$  (amplitude correction factor) and  $E_{0T}$  (calibration of the energy scale) should be used as floating parameters for each atomic edge under consideration.

**7.2.5.1. Mean Free Path, core-hole lifetime, experimental resolution,  $S_0^2$ .** The current strategy for MS calculations is based on the use of an accurate complex energy-dependent self-energy function which automatically accounts for the inelastic mean-free-path  $\lambda(E)$  (MFP) of the excited photoelectron. This also avoids the introduction of a large number of parameters. In this case, it is assumed that the inelastic losses are correctly reproduced by the model potential and no refinement of the MFP curve is performed. The program then accepts a couple of MINUIT parameters accounting for the core-hole lifetime (width  $\Gamma_h$ ) and the experimental resolution ( $\sigma_{res}$ ) expressed in eV, for each XAS measurement under consideration (number of edges  $N_{XAS} > 1$ ). The same parameters can also be assigned to different atomic edge XAS spectra.

The core-hole lifetime can be usually included as a component of the imaginary part of the potential. When a complex potential is used and the correct core-hole energy width  $\Gamma_h$  has been already considered, the MINUIT parameter should be set to zero. The experimental resolution depends on the experimental set-up used for the measurement under consideration and is usually known with sufficient accuracy. A single parameter  $\sigma_{res}$  defining the standard deviation (half-width) of the energy resolution function is used and can be slightly refined around the estimated value. Usually, it is possible to use known core-hole lifetimes and experimental resolution functions.

In some cases, one can be interested in modelling or refining the MFP curve using an optical potential which does not account for the inelastic losses. The program then accepts two sets of five MINUIT parameters (for each edge): 3 parameters describing the inelastic mean-free-path (MFP) plus 2 which take account of the core-hole lifetime and the



experimental resolution (eV). The first set of 5 parameter is applied only to the first signal (first shell), the second one to the other signals. In this way, the program is rather flexible in defining the MFP although it is always possible to define the same MFP (assignment of the same parameters for the two set) to all of the signals and edges. Of course, it is possible to FIX all of the parameters or try to refine the MFP curve. The 3 parameters of the MFP  $\lambda(E)$  curve are EXP,  $A$ , and  $B$ :

$$\lambda = A \times k^{\text{EXP}} + \frac{B}{k^4}$$

EXP is dimensionless,  $A$  and  $B$  are expressed in  $\text{\AA}^{(1+\text{EXP})}$  and  $\text{\AA}^{-3}$ , respectively. This expression can reproduce quite accurately all of the measured MFP curves.

Another important parameter is the  $S_0^2$  reduction factor, accounting for effective many-body corrections to the one-electron cross-section. Unfortunately, the multi-channel MS theory does not give a precise value for the amplitude of the principal transition, and scattered values are found in the literature. It is well-known that the overall amplitude factor  $S_0^2$  must lay in the interval 0.8–1, approximately constant as a function of the energy. Moreover, the intrinsic uncertainty in the threshold shape of the atomic cross-section affects also the actual intensity of the structural signal. Due to these sources of uncertainty in the overall amplitude of the structural signal, the program allows one to float an amplitude reduction factor  $S_0^2$ , which does not depend on the energy, in a convenient region.

*7.2.5.2. Energy scale of theoretical signal  $E_{0T}$ .* Another MINUIT parameter to be defined in `fittheo` is the difference between the energy scales of the experiment and of the calculation ( $E_{0T}$ ). The uncertainty of this parameter is due to many practical and theoretical reasons. In fact, for example, it is well-known that experimental spectra recorded in different international facilities show tiny differences in the energy calibration and, on the other hand, that it is difficult to define a precise Fermi energy for the MS calculations performed in the muffin-tin scheme. In particular, the zero of the theoretical energy scale is usually assigned to the muffin-tin interstitial potential energy level, a quantity that is generally difficult to relate precisely to the Fermi energy. Following these arguments, `fittheo` is able to float the  $E_{0T}$  parameter for each atomic edge under consideration in order to achieve the best-fit in the EXAFS region. The value of  $E_{0T}$  is usually in a range of 4–5 eV around and above each edge energy  $EE$  previously determined (in some systems the appearance of bound states near the edge can increase the difference  $EE - E_{0T}$ ). It has to be noted that the high-energy region, which is usually considered in the fitting procedure, is less sensitive to tiny movements of the  $E_{0T}$  parameter.

#### *7.2.6. Presentation of final results*

After that the minimization procedure has terminated due to the convergence to a minimum of the residual function  $R$  in the parameter space, a last special call is dedicated to perform several final calculations. These calculations can be however performed any time by selecting the appropriate MINUIT command (CALL FCN). The final value of  $R$  is calculated and written on output files together with all of the values of the fitting parameters. Structural parameters are also contained in a separate file where average distances and angles and covariance matrices are reported. Distance and variances related to distant atoms for three-body and four-body configurations are also calculated. The standard deviation(s) of the experimental data can be also estimated and a plot of the result of the calculation is produced at the beginning of the minimization procedure.

The final  $R$  value can be compared with the expected residual  $E\{R\}$  previously calculated for the final evaluation about the quality of the fit. The estimated ERRdef ( $C$ ) can be used for the determination of the error bars on the structural parameters within a selected confidence level. Correlation among parameters can be studied looking at the covariance matrices and contour maps which can be calculated using the appropriate MINUIT commands (HESSE, CONTOUR). The generation of reliable contour maps and the evaluation of appropriate error bars on structural parameters is usually obtained after a good refinement of the background and a deep minimum in the parameters' space have been reached. The special MINUIT command CALL FCN 5 is used to stop the spline refinement process to the best-fit setting currently running in the minimization. This command can be given anytime during the minimization process and it is particularly useful when the background has already been refined to a high level of accuracy and a deeper minimum in the space of the structural parameters is searched. In fact, due to the limited numerical accuracy (machine-dependent), the background can slightly deviate from the optimal background shape in the parameters' space, and this option is successfully used in calculating covariance matrices among structural parameters in the vicinity of a deep minimum. This MINUIT command is generally used to calculate the Hessian matrix (command HESSE) for the evaluation of error bars on structural parameters and correlations using contour maps (CONTOUR). See also the documentation about the Estimation of statistical errors.

A rather complete graphic output is also generated, compatible with the TOPdrawer software [21], which can be viewed on most of the workstations and graphic terminals and can produce printable Postscript files. The ASCII files containing the graphical output can be easily adapted to other graphic software. The complete graphic output generated by `fittheo` is useful to check the accuracy and consistency of the results of the fitting. In particular, the final best-fit curve and its MS decomposition is presented graphically. Individual MS contributions associated with  $g_n$  peaks are compared with the experimental data in several output files. The raw absorption data are also presented with derivatives and background function. A magnified plot showing the position of the edge energy, the atomic background, and the contribution of additional edges to the absorption and to the  $\chi(k)$  signal, is produced. It is also possible to perform the Fourier Transform (FT) of model and experimental structural signals by selecting the appropriate option. These FTs are compared also with the FT of the residual  $\chi(k) - \chi_m(k)$ , allowing a direct check of the frequency content of the signals. The FTs of each MS contribution are also calculated. Another output plot file is produced only if the mean-free-path is modelled and contains the MFP curve used in the final calculation of the model signal.

### 7.3. Use of the program

#### 7.3.1. Input files

Due to the complexity of the input cards `fittheo` is designed to work with a command (input) file (see Appendix A3 for usage under different operating systems). The `fittheo` command file must contain all the MINUIT parameters and commands plus the appropriate `fittheo` cards (see Section 7.3.3). See ref. [13] and Section 7.4 for details about MINUIT cards and commands. Basic input files are those containing the XAS data (Section 7.3.1.1) and the calculated amplitudes, phases and derivatives of  $n$ -body MS signals provided by `gnxas` (Section 7.3.1.2).

*7.3.1.1. Data files.* Input experimental data are simply the energy and absorption coefficient near the edge(s) of the system under consideration. `fitheo` then requires only two data columns: the absorption spectrum as a function of the energy, for a given number of energy points and for a given number of core-level edges, without any preliminary analysis. A certain number of EXAFS spectra related to core-level edges are read from a selected number of files (the present limit on the number of edges and files is  $NXAS=5$ ). Different edges can be stored on the same experimental file.

Different files have to be assigned in order of increasing energy. Energy values up to 200 eV are considered as measured in keV, whereas for higher values they are considered to be measured in eV. In this way the program is compatible both with eV and keV energy scale starting from a minimum energy of 200 eV.

Data must be stored in multi-column free format sequential files. Default choice for the columns is energy (1st column) and absorption (2nd column) but data can be stored on different columns by selecting the appropriate option. Lines with '#' as first character are considered comment lines and excluded from reading. Data stored using other incompatible formats should be converted to standard multi-column free format sequential files.

*7.3.1.2. MS files.* An appropriate number of theoretical input files is required in order to build the  $\chi_m(k)$  model signal. The program requires files previously calculated by using the GNXAS package (\*.DER files) or any other equivalent set of programs. These are binary files (direct access, unformatted) and contain the amplitudes, phases, their derivatives calculated using angular coordinates, and full information about the nature of the MS signal and coordinate values.

The present version of the program reads up to 60 MS files ( $NX_=60$ , this number can be largely increased) whose names have to be indicated in the command file. MS files are read as part of the cards associated with particular peaks of the  $n$ -body distribution functions. The total number of peaks of the two-body, three-body and four-body distribution functions is currently set to  $NG_=25$ . Energy scale of the MS \*.DER files is interpolated onto the experimental energy scale, according to the instantaneous  $E_{0T}$  value. When the experimental energy scale is too wide with respect to that of the theoretical signals an extrapolation of phases and amplitudes is performed whenever possible. Each MS file, containing  $\chi(k)$  or total  $\gamma(k)$  or  $\eta(k)$  signals, is manipulated according to the corresponding  $g_n$  peak. The criterium used to select the MS contributions is simply to include only those with total length less than an appropriate cut-off distance (frequency). Of course, very low-amplitude MS signals can be always neglected. This ensures that all the MS paths up to a given frequency are included to model the signal and that the difference with the experimental spectrum must contain only higher frequencies.

The quality of the fit increases including all the relevant  $g_n$  peaks which contribute to the absorption spectrum. Useful information about coordinates, energy extension and nature of the MS signals is reported while reading the \*.DER files. Amplitudes and phases of the MS signals are calculated using specified sets of initial coordinates (see `gnxas` documentation). Initial values of the coordinates should not deviate considerably from the structural values used in the fitting procedure. A check on the consistency of the coordinates associated with the MS signals is performed. The coordinates related to the MS signals can be refined up to deviations of the order of 10% from the initial values.

There are some cases where the MS signals need to be recalculated. Inspection of the graphic output can easily reveal possible distortions in the  $n$ -body  $\gamma^{(n)}$  calculated starting from the input \*.DER files. This can happen when the average coordinates used by `fitheo` are too different (more than 10%) from the initial ones, when the energy mesh used in theoretical calculations is too wide, when the signals are calculated in a too narrow energy region, or when the accuracy in the calculation of the derivatives is not sufficient. In these cases, the use of a finer and extended energy mesh and/or a slight displacement of the initial coordinates used in the `gnxas` MS calculation can help to solve the problem.

### 7.3.2. Output files

The typical output files produced by `fitheo` are described in this section including those containing the structural results, manipulation of data and final graphic presentations. Figures presented in this documentation refers to the output obtained after a typical `fitheo` run modelling silicon  $K$ -edge EXAFS data collected in 1988 at the ACO storage ring on  $\text{SiCl}_4$  gas-phase molecules. The command file used to obtain text output and figures is the first one contained in Section 7.4. Only the final results of the minimization (including the  $[1s2p]$  double-electron excitation contribution) are presented.

**7.3.2.1. Parameters files.** The final values of the parameters used in the fitting procedure, including the FIXED ones are written in an output sequential file (UNIT=7). The name of the output file is simply the three-characters name ('ABC', for example) indicated in card 4 of the command file with an extension ('ABC'.OUT).

This output file contains:

- 1) All the information of the pre-analysis such as EE (indicated as EDGE), and (if derivatives are calculated) SLOPE ( $b$ ), Jump, EWL, WL (indicated as hwl) for each energy edge under consideration (see Section 7.2.1);
- 2) spline mode and spline coefficients (when refined during the fitting procedure);
- 3) the name(s) of the data file(s);
- 4) fitting conditions such as: the number of fitting points, the total number of free parameters used, energy range used and weight, date of the fitting;
- 5) the final value of the residual function  $R$ ;
- 6) statistical information such as: standard deviation of experimental data and weight for each data file, expected  $R$  value (if selected in command file), suggested ERRdef for the calculation of the error bars on structural parameters (for more information see statistical analysis in Section 7.2.3);
- 7) the values of all the output parameters of MINUIT routine defined as  $X(N)$ , where  $N$  runs over the label numbers assigned to the MINUIT cards.

This is the output `C1n.out` obtained using the command file described in Section 7.4 after a 30 s run on a Compaq-Digital workstation working under UNIX operating system.

```
DATE: 5/ 9/2000
EDGE= 1841.233
SLOPE= 2.942756
EWL= 1842.059
hwl= 4.341204
JUMP= 0.3730208
Spline mode (user preferences): NFIT1= 3
Spline coefficients:
3.842692781445830E-002
```

```

3.262797959619736E-002
5.947428213873400E-003
FILE: /home/dicicco/gnxas/test/fittheo/./dat/sic14.
FIT: 360 points, 12 parameters, 1 g2, 1 g3, 0 g4
Energy range 1895.0 2300.0 weight ** 1.0
      R value = .72757E-05
St. Dev. of exp. data (aut. determined) =.55765E-03
Normalized St. Dev. of exp. data =.14950E-02 edge n.1
Expected R value for given weight and St. Dev.=.46868E-05
Suggested ERROR DEF (95%) for 12 free parameters ->>>.27373E-06<<<
>>> F-test OK: R value decreased of a statistical significative value
>>> (within 5% out of random variation)
>>> Minimum improvement (5%): 1.0185227
>>> Observed improvement : 3.6714149 for
>>> 4 added parameters over 348 degrees of freedom (n2)
      X( 7)= 1969.8553686
      X( 8)= 0.0164033
      X(10)= 36.5947812
      X(13)= 0.6991155
      X(15)= 2.0224181
      X(16)= 0.0014970
      X(31)= 112.1887040
      X(32)= 29.6929142
      X(33)= -0.6204787
      X(34)= 0.2715416
      X(51)= 1847.8686789
      X(52)= 4.0000000
      X(53)= 0.9151182

```

Another important output file, GN'ABC'.OUT, is used to report results on structural parameters (UNIT=60). This file is produced only if IOUT (see next subsection) greater or equal to 1 has been selected (IOUT>0 option). It contains the complete structural information associated with two-body, three-body and four-body signals introduced in the fitting procedure. Coordination numbers or actual multiplicities, distances and angular coordinates, covariance matrices and cumulants or other parameters of the distributions are listed.

Distances and variances associated with the third interatomic distance (long-bond) of the triangular configuration are also calculated. Distance and variances related to the other three longest distances in a four-body configuration are calculated up to second-order corrections.

This is the output gnCl.n.out obtained using the command file reported in Section 7.4.

```

... two-body peak n.          1
Signal n.  1 CN = 4.00000
Signal n.  2 CN = 4.00000
Signal n.  3 CN = 4.00000
R = 2.0224181 sigR**2 = 0.0014970
... three-body peak n.       1
Signal n.  1 CN = 12.00000
Signal n.  2 CN = 12.00000
Signal n.  3 CN = 12.00000
Signal n.  4 CN = 24.00000
Signal n.  5 CN = 12.00000
Signal n.  6 CN = 24.00000
Signal n.  7 CN = 12.00000
Signal n.  8 CN = 24.00000

```

```

R1 = 2.0224181      R2 = 2.0224181      theta = 112.18870
                0.0014970          0.2715416          -0.6204787
M =                0.0014970          -0.6204787
                                29.6929142

... third distance associated with the three-body peak

R = 3.3520730      sigR**2 = 0.0055831

```

Another output produced under the same conditions of the GN'ABC'.OUT is the grr'ABC'.IN file, which is used as input file for the grrrec program which reconstructs the pair distribution function.

### 7.3.2.2. Graphic output.

A detailed graphic output is produced by fitheo. All the plot files are two-column sequential files compatible with the TOPdrawer graphic software. For details about this software see ref. [21]. The latest releases of TOPdrawer are compatible with several operating systems such as Compaq-Digital OpenVMS, Digital UNIX (OSF), HP UNIX, Linux. Color plots can be viewed on workstations and personal computers including also graphic terminals and emulators. High-quality Postscript graphic files can be produced and printed on commercial laser printers. The output graphic files of fitheo can be easily converted to be readable for other available graphic software.

Partial or complete graphic output can be produced according to five different options selected by the value of the IOU parameter in the range 0–4 (see below). The output files are designed both for single-edge and multiple-edge data-analysis. In this last case, plots related to various files or edges are grouped in the same figure or in different plots contained in the same file.

There are some differences between successive versions of the TOPdrawer software [21]. The IOU parameter is also used to enhance the compatibility and the graphic performances with different TOPdrawer versions. Using intermediate TOPdrawer releases (1989–1996) the IOU parameter must be initially set to 5 (advanced graphic option) and then a second card is used to select partial or complete graphic output (actual IOU value). Recent TOPdrawer versions are compatible with color output. When IOU is initially set to 6 a color output is produced and different line intensities in various files are used for better graphic presentation of the results. Color graphics are compatible with several X-windows workstations, personal computers and terminal emulators. Line intensities are visible in the PostScript files. Color plots are also produced using IOU=7 for the last available TOPdrawer version (see ref. [21]), portable under Linux and Unix operating systems.

These files are produced for any value of IOU (IOU=0 option):

P1'ABC'.DAT (UNIT=83): double-plot file that contains data, pre-analysis (including derivative when selected) and background modelling of the data in actual photon energy scale E(eV). This plot is repeated for each data file used for the analysis.

PA'ABC'.DAT (UNIT=84): magnified plot which contains absorption data, background modelling (vs. E(eV)) and the non-structural signal obtained subtracting the  $\chi_m(E) \times \alpha_0(E)$  term. A second plot contained in this file shows the near-edge structures indicating the position of the energy zero  $E_{0T}$  and the total elastic

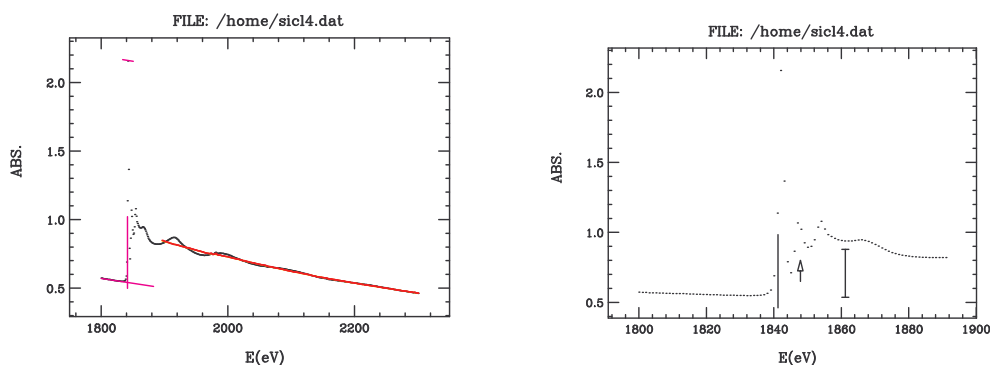


Figure 3. First two plots contained in the output file paCln.dat obtained using the command file described in Section 7.4

channel amplitude  $S_0^2 \times J$ . These plots are repeated for each data file used for the analysis.

PE'ABC'.DAT (UNIT=86): plot file that shows experimental  $\chi(k)$ , model  $\chi_m(k)^{ESP}$  and the residual (shifted) in  $k$ -space ( $\text{\AA}^{-1}$ ). When multiple-edge fitting is performed, this plot is composed of several panels related to the various edges (see ref. [12]). Date of the fitting and residual value are reported.

PT'ABC'.DAT (UNIT=90): complete plot file that contains all the contributions  $\gamma_n(k)$  associated with the  $g_n$  peaks included in the minimization, their sum  $\chi_m(k)$  (total model signal), the experimental  $\chi(k)$  and the residual (vs.  $k(\text{\AA}^{-1})$ ). Date of the fitting and residual value are reported. When multiple-edge fitting is performed, this plot is composed of several panels related to the various edges. Continuation of the  $\gamma_n(k, J)$  signals due to a selected edge  $J$  are plotted as dashed curves on the corresponding  $\gamma_n(k, K)$  signal of another edge  $K$ . The actual weight due to the ratio of the two atomic cross-section  $\alpha_0(E, J)/\alpha_0(E, K)$  is used to show the actual amplitude of the superimposed signal. Date of the fitting and residual value are reported.

FILE: /home/sicl4.dat

EDGE= 1841.2	EWL= 1842.1
SLOPE= 2.943	HWL= 4.341
JUMP= 0.373	

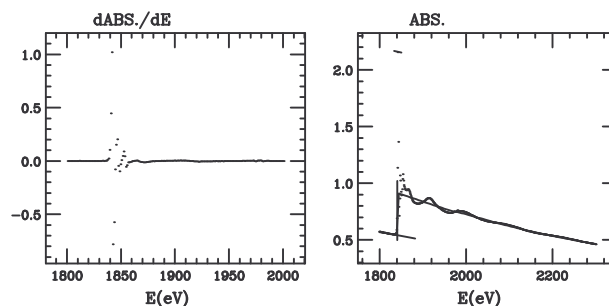


Figure 2. Graphic output p1Cln.dat obtained using the command file described in Section 7.4

Moreover, the Fourier Transform (FT) of experimental, model and residual curves can be calculated in a selected wave-vector range. If the FT option has been selected another file is produced:

PF'ABC'.DAT (UNIT=91): plot file that shows the Fourier Transform of  $\chi(k)$ ,  $\chi_m(k)$  and of the residual in  $R$ -space. Date of the fitting and residual value are reported.

If IOUT is different from zero (for example: IOUT=1): other 2 figures can be produced:

PX'ABC'.DAT (UNIT=92): this plot file contains all of the MS  $\chi_m$  or  $\gamma_n(k)$  contributions, their sum  $\chi_m(k)$  (total model signal) compared with  $\chi(k)$ , and the residual (vs.  $k$  ( $\text{\AA}^{-1}$ )). This plot is produced only when the number of MS signals exceeds that of the  $g_n$  peaks used in the fitting procedure. When multiple-edge fitting is performed, this plot is composed of several panels related to the various edges.

PA'ABC'.DAT (UNIT=84): an additional plot is added to this file if additional edges are included in the total background function. This plot shows the comparison between the experimental data  $\chi(k)$  obtained without inclusion of the  $\alpha_{exc}$  term due to additional absorption channels, and the total model signal  $\chi_m(k) + \alpha_{exc}(k)/\alpha_0(k)$ , in  $k$ -space ( $\text{\AA}^{-1}$ ). The residual curve (shifted) is compared with the shape of the additional edges, normalized to the atomic absorption of the main edge  $\alpha_{exc}(k)/\alpha_0(k)$ . The scope of this figure is to highlight possible contributions of additional edges.

If IOUT is equal or greater than 2 (for example, IOUT=2) and if the Fourier transform (FT) option has been selected (and if more than one MS signal is used in the minimization) another plot file is produced:

FG'ABC'.DAT (UNIT=95): this file contains the FT of the best-fit MS  $\gamma(k)$  contributions (signals associated with  $g_n$  peaks), calculated as appropriate sum of MS signals or as irreducible terms given by the continued fraction expansion, contained in the \*.DER files and averaged using the best-fit  $g_2$ ,  $g_3$  or  $g_4$  peak distributions. If the number of MS signals exceeds the number of  $g_n$  peaks (more than one signal is used to account for the irreducible  $n$ -body contribution), a second plot containing the FT of all of the MS signals read from \*.DER files (averaged with best-fit structural parameters) is produced.

If IOUT is equal or greater than 3 (for example, IOUT=3):

PA'ABC'.DAT (UNIT=84): another figure is added to the PA'ABC'.DAT file. It contains a magnified plot of experimental data, background, and non-structural signal, obtained subtracting a linear term from the post-edge absorption. This figure is very useful to detect possible anomalies of the atomic background and to *verify the smoothness* of the post-edge background obtained by `fittheo`. Dashed arrows indicate the position of suspected glitch points.

MF'ABC'.DAT (UNIT=50): plot file of the effective Mean Free Path(s) ( $\text{\AA}$ ) vs.  $k$  ( $\text{\AA}^{-1}$ ). This plot is produced only if MFP parameters are explicitly included in the fitting procedure:

If IOUT is equal to 4, some plot files are presented in a different way (IOUT=4):



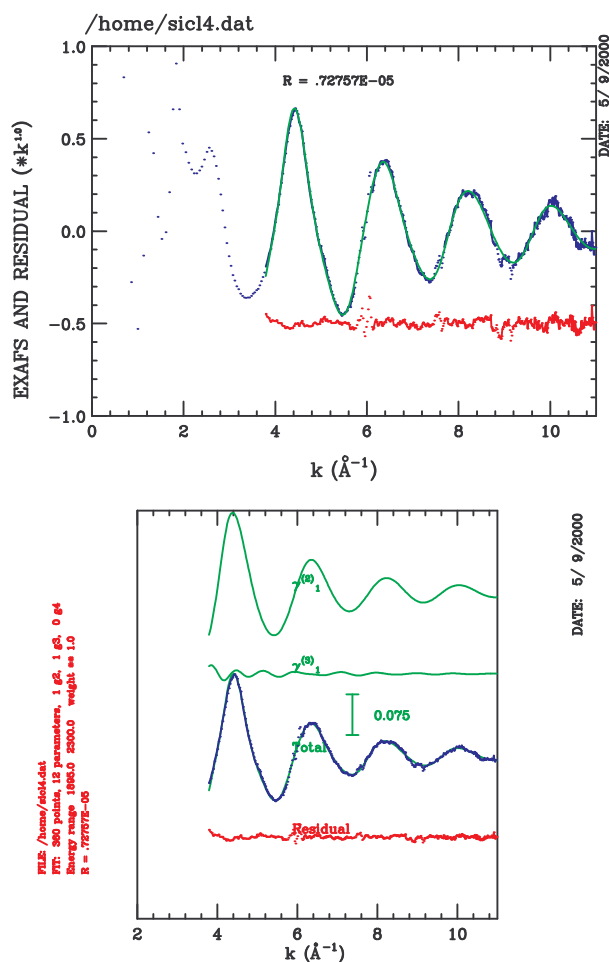


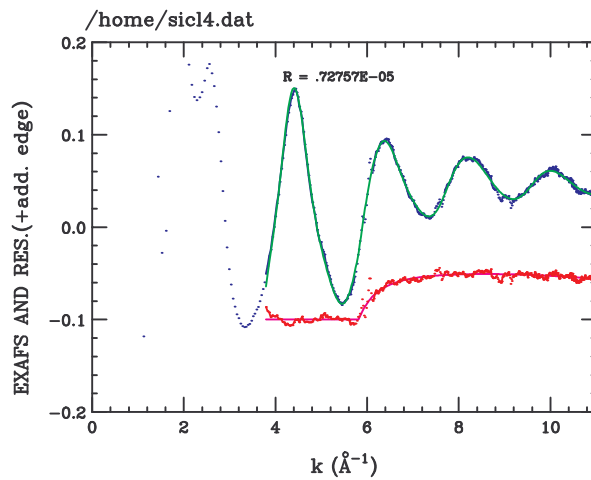
Figure 4. Output files peC1n.dat (upper) and ptC1n.dat (lower) obtained using the command file described in Section 7.4

PT'ABC'.DAT (UNIT=90) and PX'ABC'.DAT (UNIT=92) are presented with signals multiplied by  $k^{\text{ESP}}$ , for better visualization of small amplitude contributions at high wave-vector values.

7.3.2.3. *Statistical analysis output.* Expected  $R$  value  $E(\{R\})$  and the average standard deviation(s) of experimental data (noise) are written in the 'ABC'.OUT (UNIT=7) output file if the standard option is selected (IFSTO.EQ.'Y') or if this final information is explicitly requested (IFST1.EQ.'Y'). Number of data points and total number of parameters are always written on the output file 'ABC'.OUT.

The estimated ERRdef value related to the actual number of data points and floating parameters is also written in the same file, for a 95% confidence level.

When multiple-edge data are stored on different files J=1, NXAFIL, it is assumed that the quality of these measurements can be significantly different and therefore the statistical



**Figure 5.** Third plot contained in the output file `paCl1n.dat` using `IOUT≠0` in the command file reported in Section 7.4

weight of each data file

$$w_J^f = \frac{\sigma_{eff}^2}{\sigma_{expt,J}^2}$$

is computed and written on the 'ABC'.OUT file ( $w_J^f$  is always one for single-edge refinement).  $\sigma_{eff}^2$  is the average variance of the whole data set while  $\sigma_{expt,J}^2$  is the variance of measurements contained in the data file  $J$ . To increase flexibility, the weight  $w_J^f$  can be also selected by the user in the command file. For multiple-edge refinement, the standard deviations normalized to the actual absorption jump are also written in the 'ABC'.OUT file. When the  $F$ -test is applied (IFST3.EQ.'Y') to two successive minimizations performed within the same MINUIT run, the result is written in the 'ABC'.OUT file. An automatic check on the statistical significance of the inclusion of additional parameters is performed on the basis of the  $F$ -test.

Different values for the standard deviations can be used for each experimental data file, according to the noise of each measurement (for a single-edge measurement, there is only one average standard deviation). They can be inserted as input values in the command file (IFST.EQ.'N') or estimated by the program itself (IFST.EQ.'Y'). A standard choice (IFST0.EQ.'Y') is to compute the standard deviations from fitting seventh-degree polynomials in successive regions, each one including 50 energy points. Different choices can be made according to the particular spectrum under consideration (IFST0.EQ.'N').

Details of the calculations related to the statistical analysis are always written on the screen, or in the default output file (UNIT=6) during the pre-analysis `fitheo` run. In particular, a table of the energy points corresponding to deviations of the signals well above the noise level is visualized for successive identification of glitches, spikes or other evident distortions of the absorption spectra under consideration.

A specific output file containing statistical information can be generated using the option IFST2.EQ.'Y'.

PS'ABC'.DAT (UNIT=98): it is a plot file written during the pre-analysis of the experimental data. The standard deviations  $\sigma_{expt,J}^2$  of the experimental data vs.  $E$  (eV) are

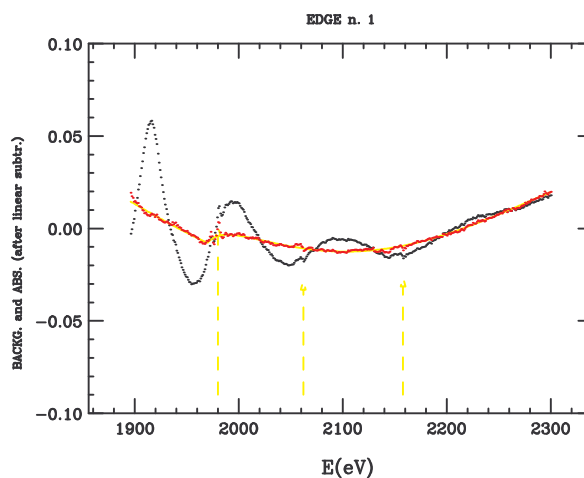


Figure 6. Plot contained in the output file paCln.dat using IOU $\geq$  3 in the command file reported in 7.4

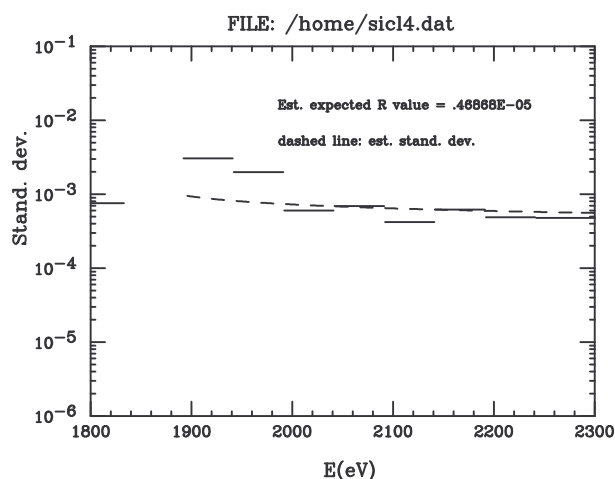


Figure 7. Plot contained in the output file psCln.dat using the command file reported in 7.4

plotted in logarithmic scale for each data file J. Average standard deviation is estimated in successive energy regions containing a selected number of points (standard choice: 50). A simple power-law expression for the standard deviation (corresponding to the  $k^w$  weighting) is also drawn on the same scale (dashed):

$$\sigma^2(k_i) = (\sigma_{expt}^2) \times \frac{k_{max}^w}{k_i^w}.$$

When the standard deviations  $\sigma_{expt,J}^2$  are given as input in the command file (IFST.EQ.'N'), this plot contains only this last curve. If standard deviations are automatically evaluated (IFST.EQ.'Y') the plot file PS'ABC'.DAT contains the comparison between experimental data and successive polynomial fits (dashed) used in the evaluation of the standard deviations for each data file.

7.3.2.4. *Warnings.* A file FITHEO.WRN (or in some fitheo versions 'ABC'.WRN) is produced anytime warning messages appears on the screen, window or in the LOG file. It can contain various messages which can be related to errors in the assignments of the input \*.DER files or problems occurred during minimization. It also contains the list of suspected glitch points detected during the statistical analysis of the absorption data.

### 7.3.3. *Cards of the command file*

In the following we describe briefly the fitheo input cards, indicating the proper format when the free format is not used. The first lines of the COMMAND file contains always the definition of the parameters used by MINUIT for the minimization.

→ fitheo cards appear after MINUIT parameters' definition. Each MINUIT parameter is defined by a line reporting number, name, initial value, initial step, minimum and maximum value for the minimization range.

WARNING: These lines are read with a format, each field is 10 character long. Make sure you have not typed a "tab" character!!!

#### 1. NXAS (INTEGER\*1)

Number of EXAFS spectra to be considered in a single fitheo run. Spectra must be stored in multi-columns, free-format sequential files. The default choice is column 1 for the energy  $E$  and 2 for absorption  $\alpha$ . If different columns are used NXAS must be typed as a negative number and the two column numbers for  $E$  and  $\alpha$  follows in a successive line. NXAS will be automatically converted to the actual positive number.

#### 1.1 NXAFIL (INTEGER\*1)

This card must be added only if NXAS is greater than 1 and indicates the number of files containing the above-mentioned NXAS EXAFS spectra. NXAFIL is assumed to be always less or equal to NXAS. Energy scale of the files should be in ascending order. Each filename must be defined, therefore NXAFIL cards 2. follows the 1.1 one.

#### 2. XASFILE (character\*70) (L)

The complete filename of the file which contains the raw x-ray absorption data under consideration. If more than one data file is used (NXAFIL>1) a series of L=1, NXAFIL of these cards must be defined.

#### 3. COMNAM (character\*3)

Set of three characters to be used as common name for all output files.

#### 4A. EE1, E1 (K)

The cards 4A., 4B. are used for the pre-analysis of absorption data. If more than one edge (NXAS>1) is analysed, there must be successive sets of card 4A., 4B. associated with the various edges (K=1, NXAS). In card 4A., EE1 and E1 are the lower and upper energy limits used to fit the pre-edge (they have to be inside the energy limits of the actual data). They must be two energy values between which the pre-edge has a smooth linear behaviour. All the points between them in the data file (extremes included) are used for a linear fit of the pre-edge. Spikes should be excluded from the interval EE1-E1, as well as the beginning of the rise of the absorption edge.

#### 4B. E2, TOLL (K)

E2 is defined as the maximum energy for derivative calculation and TOLL is a tolerance parameter. The two-point derivative of the spectrum is calculated from the first energy point to the energy E2. The tolerance parameter TOLL is a real number from 0 to 1 (usually 0.4-0.6, for noisy spectra 0.9-0.99). If E2 is less or equal to 0 the derivative

is not calculated and a successive card 4B.1 is read. No final plots on derivative and information about energy edge extraction are produced in this case. See Section 7.2.1 for further information.

#### 4B.1 EE (K)

The edge energy EE in eV must be typed in if E2 is less or equal to zero (see card 4B). This allows one to fix the EE energy using a selected value.

#### 5. CH6 (character\*1)

If a positive answer Yes or Si is given (CH6.EQ.'S' or CH6.EQ.'Y') the program performs a reduction of the effective number of energy points. This is done by interpolating the original data over a new mesh equally spaced in the wave-vector  $k$ -space. This can reduce considerably the computing time. The new energy mesh is defined by the card 5.1 that follows. This option can be selected only for single-edge refinement (NXAS=1).

#### 5.1 Dk

This card appears only if CH6.EQ.'Y' or 'S' (see card 5). Dk is a real number ( $1/\text{\AA}$ ) which defines the constant step in the wave-vector  $k$ -scale ( $k \propto \sqrt{E}$ ). If the new number of energy points is greater than that of the original data a warning message is generated suggesting the use of higher values of Dk.

#### 6A. NSPL(K)

The cards 6A, 6B, 6C, 6D are used to define the post-edge background. If more than one edge (NXAS>1) is analysed, there must be NXAS successive sets of cards 6A, 6B, 6C, and 6D associated with the various edges, in order of increasing photon energy (K=1,NXAS). NSPL is the number of spline intervals used in the modelling of the post-edge background for the edge under consideration. Values of 1–3 are commonly used, depending on the number of energy points.

#### 6B. EI(1),...,EI(NSPL+1) (K)

The EI( $i$ ) vector defines the energy intervals (eV) of the splines for the edge under consideration. The first will be used in the energy range EI(1)–EI(2), the second one in EI(2)–EI(3) and so on. The first and the last energy values EI(1) and EI(NSPL+1) represents also the total energy range used in the refinement process. Therefore these are *crucial quantities* of the fitting procedure. Specified regions of the energy intervals can be successively considered with weight 0 (see card 8).

#### 6C. NSP(1),...,NSP(NSPL) (K)

The NSP( $i$ ) vector contains the polynomial degrees of each spline  $i$  for the edge  $K$ . These parameters, together with the NSPL value, determine the total number of spline coefficients used in the background modelling (see Section 7.2.2).

#### 6D. CH2 (character\*1)

If a positive answer Yes or Si is given (CH2.EQ.'S' or CH2.EQ.'Y') the program uses a constraint for the ratio of the first two spline coefficients. This implies the use of an Hydrogen-like background around the edge and determines a reduction of one parameter in the spline parameterization. When the original data departs from this behaviour this option should not be used. The special option CH2.EQ.'J' or means that the Jump of the edge (K) under consideration is kept fixed to the input value accepted in the successive card 6D.1. The option CH2.EQ.'D' implies the application of the above-mentioned constraint at the edge for fixed Jump.

6D.1 JUMP (K) This card appears only if CH2.EQ.'J' or CH2.EQ.'D' and defines the jump for the edge under consideration, which will be kept fixed during the refinement.

## 7. NEDGE

The integer NEDGE defines the number of additional jumps (secondary edges) to be added to the background at certain energies (see next cards). The successive 7.1 card is repeated NEDGE times.

### 7.1 T\_ED, N\_ED, N\_J2, N\_DE2

A number of NEDGE cards of this type are requested (see card 7, if NEDGE=0 no cards like this should appear). T\_ED is a character\*1 read in free format. It is followed by three (or more) integer numbers N\_ED, N\_J2, N\_DE2, ... corresponding to MINUIT parameters. T\_ED selects the functional shape for the secondary edge. T\_ED.EQ.'A', 'E', 'P', 'F', and 'L' correspond to the arctangent-like, smoothed step, smoothed slope changes, or Lorentzian shapes, respectively. These functional forms depend upon three (or more) parameters. N\_ED is the (MINUIT) parameter number for the secondary edge energy (eV), N\_J2 is the parameter number for the jump (height), and N\_DE2 is the parameter number for the onset energy interval (eV) (width for the Lorentzian case).

## 8. NINTE

NINTE is the number of intervals with null weight used in the fitting procedure. These intervals allow to exclude particular regions of the absorption spectra that can not be modelled (glitches, additional edge thresholds, distorted and noisy regions...). A series of NINTE 8.1 cards follows this one.

### 8.1 ENUL(i, 1), ENUL(i, 2)

This card has to be repeated NINTE times and indicates the intervals of energy excluded from the fit. ENUL(i, 1)-ENUL(i, 2) represents the  $i^{\text{th}}$  interval with 0 weight in the minimization process ( $i=1, \dots, \text{NINTE}$ ).

## 9. PES, RMAX

This card controls the weight of the sum appearing in the residual  $\chi^2$ -like function  $R$  and is an important factor in the minimization process. The weight is proportional to  $k^{\text{PES}}$  in each EXAFS spectrum under consideration; PES=0. gives equal weight to each point. RMAX is a particular wavelength with a weight zero in the residual (see Section 7.2.3,  $w=\text{PES}$ ). In general, RMAX=0. should be used.

## 10. CH5 (character\*1)

If a positive answer Yes or Si is given (CH5.EQ.'S' or CH5.EQ.'Y') the program assumes the default setting of the spline generation routine (*i.e.* NFIT1=3, NINFO=10, see card 10.1).

### 10.1 NFIT1, NINFO

This card does not appear if CH5.eq.'Y' or 'S'. NFIT1 is the number of calls performed without spline refining during the minimization with  $\chi_n$  signals. The NFIT1=3 option is the standard one. NINFO controls the flux of information about spline generation. The standard option NINFO=10 means that the spline coefficients are visualized every 10 calls. For very long trials we suggest greater values for NINFO. A convergence of the spline coefficients indicates a convergence to a minimum of the fitting procedure. NFIT1 cannot be less than 3, while NFIT1=4 correspond to

specialized calculations. NFIT1=4 means that no final refining will be done and the background minimization will be stopped at the last useful MINUIT call (not when the MINUIT command CALL FCN is used). The special choice NFIT1=0 and NINF0=0 means that spline coefficients defining background are MINUIT parameters. Therefore if a number NTOT of coefficient (easily calculated and indicated by the program) has to be introduced in the MINUIT cards starting from the first (X(1)) parameter of the LIST (see Section 7.2.2). This choice can be used when one wants to FIX a specified background shape. The convergence is in any case slower.

#### 11. ICLM

ICLM controls the use of the Mean Free Path (MFP) damping factor of the MS contribution. When ICLM=1 3 coefficients for the MFP curve and 2 for core-hole lifetime and experimental resolution are requested as MINUIT parameters (see Section 7.2.5). Other values for ICLM (ICLM  $\neq$  1) are used when only the core-hole lifetime and experimental resolution factors are taken into account, assuming that inelastic losses (MFP) are correctly reproduced by the imaginary part of the potential in phagen. The value of ICLM thus determines different input parameters in the subsequent card 12. If ICLM=1 two (12,12.1) cards follow.

#### 12. NPCLM(1,1), NPCLM(2,1)...NPCLM(4,1), NEXPR (if ICLM=1) (K)

NPCLM(4,1), NEXPR (if ICLM $\neq$  1) (K)

NPCLM is a vector (matrix) containing the Mean Free Path parameter used in the fitting procedure. NEXPR indicates the parameter for the experimental resolution. If ICLM=1 (see card 11) the indication of five MINUIT parameter numbers is needed: EXP, A, B, Gamma ( $\Gamma_h$ ), expres  $\sigma_{res}$  (see Section 7.2.5) for the first MS signal. If ICLM $\neq$  1 only the last two ones are needed:  $\Gamma_h$  and experimental resolution  $\sigma_{res}$  (used for all of the MS signals). When a multiple-edge fit is attempted, this card is repeated NXAS times (K=1, NXAS), in order to take into account different core-hole lifetimes and experimental resolution factors.

#### 12.1 NPCLM(1,2), ...NPCLM(4,2)

This card appears only if ICLM=1, with the same format of card 12. The MFP parameter numbers defined here are applied to all of the MS signals apart from the first one. Of course, the same MFP function can be still applied to all of the MS signals by repeating in card 12.1 the parameter numbers already indicated in card 12. When a multiple-edge fit is attempted, this card is repeated NXAS times (K=1, NXAS).

#### 13. NCORD

NCORD is the number of different coordination numbers (CNs) used in the simulation (not less than 1). The same CN can be associated with several  $g_2$ ,  $g_3$  or  $g_4$  peaks, therefore it is not necessary to use many different CNs. Indeed, the analysis of structures with known CNs can be facilitated by introducing effective multiplicity factors for each MS signal which take account of the CNs without introducing useless MINUIT parameters (see cards related to MS signals). Particular constraints can be defined resulting in a reduced number of independent CNs. An useful constraint is used to define and fix the total number of atoms related to a couple of different shells. This is accomplished through a special usage of the next 13.1 card, which is repeated NCORD times in the command file. Another important constraint can be

used to perform structural refinement of highly disordered systems (see Section 9 for details).

#### 13.1 NPCO(*i*)

NPCO(*i*) is the vector which contains the MINUIT parameters numbers for each independent coordination number (CN). The command file must contain NCORD (see card 13) cards of this type ( $i=1, \text{NCORD}$ ). The coordination numbers are then numbered according to their storage in the NPCO vector. In some special cases the relationship between MINUIT parameters and actual coordination numbers used by the program can be changed, allowing for the introduction of useful physical constraints. For example, by selecting NPCO(*i*)=0 for a chosen coordination number *i*, the programs will read a second time NPCO(*i*), assuming this time that the *i*<sup>th</sup> CN is given by  $N \times x$  (where *N* is the value of MINUIT parameter NPCO(*i*-1) and *x* is the value of MINUIT parameter NPCO(*i*)), while the (*i*-1)-th CN is given by  $N \times (1-x)$ . Therefore, when *x* is floated between 0 and 1 and *N* corresponds to a selected value, it is possible to vary the CNs of two shells of atoms keeping fixed the total number of neighbors *N*.

#### 14. NG2

NG2 is the number of different  $g_2$  peaks included in the simulation. The peaks are defined as particular two-body structures used in calculation of MS signals (see Section 7.2.4). Follows a total of NG2 14.1+14.2+14.3 cards defining the parameters associated with each peak.

##### 14.1 FGG(*i*)

FGG is a character\*1 read in free format. This card defines the type of  $g_2$  peak used for the configurational average of the MS signals. Three models are currently used: the standard Gaussian 'G', 'E' ( $\Gamma$ -like distribution), and 'C' (Cumulant expansion). The latter two ones are used for deviations from simple Gaussian distributions. The choice of the type of the distribution influences the reading of the successive 14.2 card, because the various models need definition of a different number of parameters. The command file must contain NG2 cards of this type ( $i=1, \text{NG2}$ ). Each 14.1 card is followed by one set of 14.2+14.3 cards.

##### 14.2 NR(*i*), NSIG(*i*), (NBET, NKUR)

This set of integers (read in free format) are the MINUIT parameters numbers defining the particular *i*<sup>th</sup>  $g_2$  peak. If a Gaussian distribution is used (FGG(*i*)='G') the program reads only NR(*i*) and NSIG(*i*). They represent the MINUIT parameters for the average values of distance *R* and variance  $\sigma^2$  associated with the  $g_2$  peak. When a different model distribution is used, the third parameter NBET defines the MINUIT parameter associated with the skewness of the distribution (dimensionless  $\beta$  parameter). When the cumulant expansion is used a fourth parameter (NKUR) defines the kurtosis of the distribution. These parameters can be anyway FIXed using MINUIT commands, or floated to span from Gaussian to Non-Gaussian distributions. The command file must contain NG2 cards of this type ( $i=1, \text{NG2}$ ). Each 14.2 card is followed by one 14.3 card.

##### 14.3 NPATH(*i*)

NPATH(*i*) defines the number of MS signals used in the simulation. In fact, it must correspond to the number of \*.DER files to be read. These can contain



either irreducible  $n$ -body  $\gamma^{(n)}$  or  $\chi_m$  MS amplitudes and phases. Follows NPATH(i) 14.4+14.5 cards for each peak i defining some characteristics of the MS signals.

#### 14.4 FILEDER (character\*70)

The complete filename which contains amplitudes, phases and derivatives of the particular MS signal considered has to be typed in. Only \*.DER binary files are accepted (see gn<sub>x</sub>as documentation).

#### 14.5 NC(j), AD(j)(, NXEDGE(j))

NC(j) indicates the proper coordination number (CN) as defined by the cards 13-13.1. The NC(j) values are in the range 1, NCORD and give to each signal the appropriate CN (CN is the MINUIT parameter X(NPCO(NC(J))). AD(j) is a correction factor for the multiplicity of the  $j^{\text{th}}$  MS signal. AD=1.0 when the CN corresponds to the correct coordination number associated with the signal. The AD correction factor is also useful to define the correct CN or multiplicity without introducing several CN MINUIT parameters. When coordination numbers are known it is possible to define all of the multiplicities of the MS signals, defined as CN×AD through one CN only. This can be achieved, for example, by defining the first coordination number X(NPCO(1))=1 and AD factors equal to the actual multiplicity of each MS signal. The third parameter, NXEDGE, defines the edge number (edges are numbered from 1 to NXAS) and must be added only when a multiple-edge fitting is performed. Different signals related to the same  $g_2$  peak are used to simulate different absorption edges. There are NPATH(i) cards of this type (14.5).

#### 15. NG3

NG3 is the number of different  $g_3$  peaks included in the simulation. The peaks are defined as particular three-body structures used in calculation of MS signals (see Section 7.2.4). Follows a number of NG3 15.1+15.2+15.3 cards defining the parameters associated with each peak.

##### 15.1 FGG

FGG is a character\*1 read in free format. This card defines the type of  $g_3$  peak used for the configurational average of the MS signals. A standard Gaussian function 'G' and the cumulant expansion 'C' can be selected. The command file must contain NG3 cards of this type (i=1, NG3). Each 15.1 card is followed by one set of 15.2+15.3 cards.

##### 15.2 NR1, NR2, NTHE, NSIG1, NRHO12, NRHO1T, NSIG2, NRHO2T, NSIGT (i)

These integers are the MINUIT parameters numbers for the particular  $i^{\text{th}}$   $g_3$  peak. The first three parameters represent the average values of the two distances R1 and R2 and of the angle  $\theta$  (THE) defining the  $i^{\text{th}}$  triangular structure included in the fit. The last six numbers define the covariance matrix and the meaning of each parameter is illustrated in Section 7.2.4. Here we recall that: the  $\rho$  (RHO) parameters take into account bond-bond (RHO12) and bond-angle (RHO1T, RHO2T) correlations and are defined in the range [-1,1];  $\sigma_{R1}^2$  (SIG1) and  $\sigma_{R2}^2$  (SIG2) parameters are the bond variances to be used also for the corresponding peaks of the  $g_2$ ;  $\sigma_{\theta}^2$  (SIGT) is the variance (deg<sup>2</sup>) of the angle  $\theta$  (THE). The COMMAND file must contain NG3 cards (see card 15) of this type (i=1, NG3) When the cumulant expansion is selected, other six dimensionless parameters related to the third and fourth cumulants of the distribution of the bond and angular coordinates have to be added to the list: NBET\_1,

NKUR\_1, NBET\_2, NKUR\_2, NBET\_T, NKUR\_T. Of course, all of these parameters can be FIXED (to zero) using MINUIT commands. Each 15.2 card is followed by one 15.3 card.

### 15.3 NPATH(i)

NPATH(i) defines the number of MS signals used in the simulation for calculating the contribution of the  $i^{\text{th}}$   $g_n$  peak to the XAS structural signal ( $g_3$  peak). It must correspond to the number of \*.DER files to be read successively. Follows NPATH(i) 15.4+15.5 cards for each peak i defining the characteristics of the MS signals.

### 15.4 FILEDER (character\*70)

The complete filename which contains amplitudes, phases and derivatives of the particular MS signal considered has to be typed in. Only \*.DER binary files are accepted (see gnexas documentation).

### 15.5 NC(j),AD(j),NXEDGE(j)

NC(j) indicates the proper multiplicity of the XAS signal as defined by the cards 13-13.1. The NC(j) values are in the range 1,NCORD and give to each signal the appropriate multiplicity factor. AD(j) is a correction factor for the multiplicity DEG(j) of the  $j^{\text{th}}$  MS signal related to a  $g_3$  peak. AD=1.0 is when the NC(j) value corresponds to the correct multiplicity of the signal. The multiplicity DEG(j) is defined as CN×AD (being CN=X(NPCO(NC(j))), MINUIT parameter). In this way any value of DEG(j) can be selected recovering the actual multiplicity values for the particular structure considered or those assigned previously in the ab-initio gnexas calculations. The third parameter, NXEDGE, defines the edge number (edges are numbered from 1 to NXAS) and must be added only when a multiple-edge fitting is performed. Different signals related to the same  $g_3$  peak are used to reproduce the XAS spectra of different absorption edges. There are NPATH(i) cards of this type (15.5).

## 16. NG4

It is the number of different  $g_4$  peaks included in the simulation. The peaks are defined as particular four-body structures used in calculation of MS signals (see Section 7.2.4). Follows a number of NG4 16.1+16.2+16.3 cards defining the parameters associated with each peak.

### 16.1 FGG

FGG is a character\*1 read in free format. This card defines the type of  $g_4$  peak used for the configurational average of the MS signals. A standard Gaussian function 'G' and the cumulant expansion 'C' can be selected. The COMMAND file must contain NG4 cards of this type (i=1,NG4). Each 16.1 card is followed by one set of 16.2+16.3 cards.

### 16.2 NR1,NR2,NTHE,NR3,NFI,NPSI,NM(j=1,21) (i)

The indicated 27 integers are the MINUIT parameters numbers for the particular  $i^{\text{th}}$   $g_4$  peak. The first six numbers represent the average values of the three distances R1 and R2, R3 and of the angles  $\theta$ ,  $\phi$ ,  $\psi$  (THE, FI, PSI). The other 21 parameters (NM) define the covariance matrix of a  $g_4$  peak defined row by row:

```
NSIG1, NRHO12, NRHO1T, NRHO13, NRHO1F, NRHO1P,
NSIG2, NRHO2T, NRHO23, NRHO2F, NRHO2P,
NSIGT, NRHOT3, NRHOTF, NRHOTP,
NSIG3, NRHO3F, NRHO3P,
```

NSIGF, NRHOFP,  
NSIGP.

The meaning of these parameters is illustrated in Section 7.2.4 The NSIGn (n=1,3) integers define the MINUIT parameters associated with the variances of distance coordinates, the other NSIG values are associated with the angular variances (T, F, P are  $\theta$ ,  $\phi$ ,  $\psi$ , respectively).  $\rho_{ij}$  (RHO) parameters are the bond-bond and bond-angle correlations and are defined in the range [-1,1]. When the cumulant expansion is selected, other 12 dimensionless parameters related to the third and fourth cumulants of the distribution of the bond and angular coordinates have to be added to the list:

NBET\_1, NKUR\_1, NBET\_2, NKUR\_2, NBET\_T, NKUR\_T,  
NBET\_3, NKUR\_3, NBET\_F, NKUR\_F, NBET\_P, NKUR\_P.

Of course, all of these parameters can be FIXEd (to zero) using MINUIT commands. Each 16.2 card is followed by one 16.3 card. The COMMAND file must contain NG4 cards (see card 16) of this type (i=1,NG4).

#### 16.3 NPATH(i)

NPATH(i) defines the number of MS signals used in the simulation of the contribution of the  $i^{\text{th}}$   $g_4$  peak. It must correspond to the number of \*.DER files to be read. Follows NPATH(i) cards 16.4+16.5 for each peak i defining some characteristics of the MS signals.

#### 16.4 FILEDER (character\*70)

The complete filename which contains amplitudes, phases and derivatives of the particular MS signal considered have to be typed in. Only \*.DER binary files are accepted (see gn<sub>x</sub>as documentation).

#### 16.5 NC(j),AD(j)(,NXEDGE(j))

NC(j) indicates the proper multiplicity for the four-body MS signal as defined by the cards 13–13.1. The NC(j) values are in the range 1,NCORD and are used to assign to each signal the appropriate multiplicity factor. AD(j) is a correction factor for the multiplicity DEG(j) of the  $j^{\text{th}}$  MS signal related to a  $g_4$  peak. AD=1.0 when the NC(j) value corresponds to the correct multiplicity of the signal. The multiplicity factor DEG(j) is defined as CN×AD, where X(NPCO(NC(j))) is the CN MINUIT parameter. In this way any value of DEG(j) can be selected recovering the actual multiplicity values for the particular four-body signal under consideration or those already assigned in the ab-initio gn<sub>x</sub>as calculations for the model structure. The third parameter, NXEDGE, defines the edge number (edges are numbered from 1 to NXAS) and must be added only when a multiple-edge fitting is performed. Different signals related to the same  $g_4$  peak are used to simulate different absorption edges. There are NPATH(i) cards of this type (16.5).

#### 17. NET(K)

NET is the MINUIT parameter number for the  $E_{0T}$  parameter described in Section 7.2.5. The value of  $E_{0T}$  is generally in a range of 4–5 eV around the edge energy.  $E_{0T}$  is accepted in eV for values higher than 200 and in keV for values lower than 200. This card must be repeated defining  $E_{0T}$  for all of the energy edges under consideration K=1,NXAS.

18. NSOQ(K)  
NSOQ is the MINUIT parameter number for the  $S_0^2$  parameter described in Section 7.2.5. The value of  $S_0^2$  is usually in the range 0.8–1. This card must be repeated defining  $S_0^2$  for all of the energy edges under consideration K=1, NXAS.
19. CH1 (character\*1)  
If a positive answer Yes or Si is given (CH1.EQ.'S' or CH1.EQ.'Y') the program calculates a Fourier Transform (FT) with Gaussian window of the experimental, theoretical and residual spectra. If IOUT is larger than 2 (see card 20) the FT is calculated for each  $\gamma(k)$  signal and/or MS particular signal. The  $k$ -range and weight of the FT are defined by the card 19.1 that follows.
- 19.1 AK1, AK2, ESP  
This card appears only if CH1.EQ.'Y' or 'S'. AK1 and AK2 are the limits of the Fourier Transform in  $k$ -space ( $\text{\AA}^{-1}$ ). If those limits exceed the actual energy limits of the fit AK1 and AK2 are automatically corrected. The original spectra in  $k$ -space are multiplied for  $k^{\text{ESP}}$  allowing for an increasing weight for the high-energy region (usually ESP=1–3). The same exponent ESP is used in the plot the structural signal in wave-vector  $k$ -space.
20. IOUT  
This card selects the quantity and type of plot files that the program is going to produce (IOUT=0–4, see Section 7.3.2). IOUT=5,6,7 have to be used for better graphic performances using recent TOPdrawer releases (see Section 7.3.2). When IOUT=5,6,7 this 20 card must be repeated two times: the first IOUT value is used to select the output format for various versions of TOPdrawer, the second one is the actual IOUT value used by fittheo.
21. STDEV(L)  
STDEV(L) is the standard deviation of raw data contained in each experimental input file (one only for single-edge refinements) L (average statistical noise). If the first standard deviation is zero (STDEV(1)=0.0) the program automatically evaluates STDEV(L) using a fitting procedure.
- 21.1 ENMIN  
This card appears only if STDEV(1)=0.0. ENMIN defines the starting energy to evaluate the noise level. If ENMIN=0.0 the edge is assumed as starting point. For multiple-edge fits the starting energy always coincides with the edge energy.
22. IFST0 (character\*1)  
If a positive answer Yes or Si is given (IFST0.EQ.'Y') the standard choice for the statistical analysis is selected. This includes: standard 50 points estimate of variance, statistical information on output files ('ABC'.OUT), weight of different experimental files automatically calculated. If a negative answer is given (IFST0.NE.'Y') a set of additional cards 22.1–22.4 is needed.
- 22.1 LENG(L)  
This card is needed only if IFST.NE.'Y'. LENG(L) is the number of energy points considered for the calculation of the variance of the experimental data (usually 30–60) for each data file L.
- 22.2 SIGWEI(L)  
This card is needed only if IFST.NE.'Y'. It allows one to select the weight  $w_L^f$  (SIGWEI(L)=) of each experimental data file L (see Section 7.3.2.3). If  $w_L^f=0$  the

weight is automatically calculated according to the variance of each experimental data file.

### 22.3 IFST1 (character\*1)

If a positive answer Yes or Si is given (IFST1.EQ.'Y') the program writes statistical information on \*.OUT output file (for example expected residual  $R$  and suggested ERRdef values).

### 22.4 IFST4 (character\*1)

If a negative answer No is given the expected value of the residual is calculated assuming power-law shape for the variance and data points equally spaced in energy.

### 23. IFST2 (character\*1)

If a positive answer Yes or Si is given (IFST2.EQ.'Y') the program produces an output plot file containing the standard deviation  $\sqrt{\sigma_{expr}^2}$ .

### 24. IFST3 (character\*1)

If a positive answer Yes or Si is given (IFST3.EQ.'Y') the program performs the statistical  $F$ -test on selected parameters. The statistical significance of adding an additional parameter is checked by using the RELEASE par # (MINUIT command) after CALL FCN in MINUIT cards (see ref. [13]) and allowing a new search of the minimum (SIMPLEX or MIGRAD) and calculating the final results (new CALL FCN MINUIT command). The  $F$ -test is applied and statistical results are written in 'ABC'.OUT file.

→ MINUIT commands follow.

## 7.4. Examples of command files

### 1) SIMPLE SYSTEM WITH MS PATHS UP TO $\chi_6$ (molecule: $SiCl_4$ )

```

EXAFS FITTING PARAMETERS
( 1      J      .37709 .00010  0.000  1.000)  MINUIT command
( 2      a     -.05692 .00010 -1.000  1.000)  -
( 3      b      .04609 .00010 -1.000  1.000)  -
( 4      c     -.00857 .00010 -1.000  1.000)  -
 7      ED    1970.96      1.  1955.  1980.  -
 8      J1     .00000 .0001  .0000  .0350  -
10      DE     35.000      1.    5.    40.   -
13      expres  0.30      .05   .200   .700  -
14      gamma  0.0000    .10   .0000  0.500  -
15      R1     2.02147    .01   1.950  2.100  -
16      Sig2-1 .0012     .0005 .001000 .0040  -
31      the    111.7083    .1    107.   115.   -
32      Sig2-the 29.6947    .005  0.50   64.0  -
33      rho1/2-t -.54985    .005 -0.65  0.65  -
34      rho12  .19000    .005 -0.65  0.65  -
51      Eo     1847.86    .20  1835.0 1850.0  -
52      Nc1    4.000     0.00  3.9   4.1   -
53      S02    0.9558    0.02  0.75  1.20  -

1                                CARD 1
../dat/sicl4.due                 CARD 2
Cln                               CARD 3
1800.,1833.                       CARD 4A
2000.,.5                           CARD 4B
N                                   CARD 5
1                                   CARD 6A

```

1895.,2300.	CARD 6B	
4	CARD 6C	
S	CARD 6D	
1	CARD 7	
'E',7,8,10	CARD 7.1	
4	CARD 8	
1966.,1983.	CARD 8.1	
2050.,2065.	CARD 8.1	
2135.,2142.	CARD 8.1	
2156.,2161.	CARD 8.1	
1.,0.	CARD 9	
N	CARD 10	
3,30	CARD 10.1	
0	CARD 11	
14,13	CARD 12	
1	CARD 13	
52	CARD 13.1	
1	CARD 14	
'G'	CARD 14.1	
15,16	CARD 14.2	
3	CARD 14.3	
../gnxas/SC1G2X2.DER	CARD 14.4	
1,1.	CARD 14.5	
../gnxas/SC1G2X4.DER	CARD 14.4	
1,1.	CARD 14.5	
../gnxas/SC1G2X6.DER	CARD 14.4	
1,1.	CARD 14.5	
1	CARD 15	
'G'	CARD 15.1	
15,15,31,16,34,33,16,33,32	CARD 15.2	
8	CARD 15.3	
../gnxas/SC1G3P1X301.DER	CARD 15.4	
1,3.	CARD 15.5	
../gnxas/SC1G3P1X401.DER	CARD 15.4	
1,3.	CARD 15.5	
../gnxas/SC1G3P1X402.DER	CARD 15.4	
1,3.	CARD 15.5	
../gnxas/SC1G3P1X501.DER	CARD 15.4	
1,6.	CARD 15.5	
../gnxas/SC1G3P1X503.DER	CARD 15.4	
1,3.	CARD 15.5	
../gnxas/SC1G3P1X601.DER	CARD 15.4	
1,6.	CARD 15.5	
../gnxas/SC1G3P1X602.DER	CARD 15.4	
1,3.	CARD 15.5	
../gnxas/SC1G3P1X603.DER	CARD 15.4	
1,6.	CARD 15.5	
0	CARD 16	
51	CARD 17	
53	CARD 18	
Y	CARD 19	
4.0,11.0,1.	CARD 19.1	
7	CARD 20	
3	CARD 20	
0.	CARD 21	
1000.	CARD 21.1	
Y	CARD 22	
Y	CARD 23	
Y	CARD 24	
SET ERRdef	5.E-10	MINUIT command
FIX	7	-
FIX	8	-

```

FIX          10          -
FIX          16          -
SIMPLEX     150          -
(MIGRAD)
CALL FCN     3          -
RELEASE     7           -
RELEASE     8           -
RELEASE     10          -
SIMPLEX     600          -
RELEASE     16          -
SIMPLEX     300          -
SET ERRdef  2.E-04      -
SIMPLEX     1000         -
EXIT
    
```

2) A MONATOMIC CRYSTALLINE SYSTEM:  $\beta$ -TIN, MULTIPLE K+L EDGE FIT

```

                                EXAFS FITTING PARAMETERS
2          NC1          4.0      .000      3.000      5.000      MINUIT command
3          NC2          2.0      .000      1.000      3.000      -
6          Eo(L3)      3936.31   0.10     3936.0     3936.6     -
7          Eo(L2)      4163.60   0.10     4163.3     4163.9     -
8          Eo(L1)      4471.29   0.10     4470.8     4471.8     -
9          Eo(K)       29.1753   0.0005   29.155     29.195     -
12         exp(K)      6.6878    .01      5.00      7.00      -
13         exp(L)      0.6000    .01      0.30      0.70      -
14         gamma      0.0000    .01      .000      0.500     -
15         R1          3.02179   .002     2.900     3.100     -
16         Sig2-1     .01278    .001     .0050     .0200     -
17         R2          3.18756   .002     3.100     3.300     -
18         Sig2-2     .02668    .001     .0050     .0550     -
19         bet1       0.11592   .005     0.000     0.600     -
20         bet2       0.13930   .005     0.000     0.600     -
31         the        74.74     .0       70.       80.       -
32         Sig2-t1    9.1391    .005     1.00      16.0     -
33         the        93.97     .0       89.       98.       -
34         Sig2-t2    24.8263   .005     1.00      25.0     -
35         the        105.26    .0       100.      110.      -
36         Sig2-t3    15.9975   .020     1.00      16.0     -
37         rho        -.000     .00      -0.45     0.45     -
41         ED        29670.28   1.0     29670.0   29740.0   -
42         J1         .00306    0.001    0.00      0.05     -
43         DE         9.7640    1.0      3.0       50.0     -
44         ED        29277.59   1.0     29250.0   29290.0   -
45         J1         .00950    0.001    0.00      0.09     -
46         DE         19.376    1.0      3.0       50.0     -
51         S02(L3)    0.910    0.02     0.80      1.00     -
52         S02(L2)    0.930    0.02     0.80      1.00     -
53         S02(L1)    0.860    0.02     0.80      1.00     -
54         S02(K)     0.900    0.02     0.80      1.00     -

4          CARD 1
2          CARD 1.1
SNF.rt     CARD 2
SNRTK.dat  CARD 2
sft        CARD 3
3890.,3920. CARD 4A
0.,.5     CARD 4B
3936.8     CARD 4B.1
4070.,4130. CARD 4A
0.,.5     CARD 4B
    
```

4163.8	CARD 4B.1
4370.,4440.	CARD 4A
4550.,.5	CARD 4B
28970.,29070.	CARD 4A
29250.,.5	CARD 4B
N	CARD 5
2	CARD 6A
3965.,4050.,4140.	CARD 6B
4,4	CARD 6C
S	CARD 6D
2	CARD 6A
4192.,4270.,4420.	CARD 6B
4,5	CARD 6C
S	CARD 6D
2	CARD 6A
4494.,4570.,4621.	CARD 6B
4,3	CARD 6C
S	CARD 6D
3	CARD 6A
29210.,29320.,29550.,29999.	CARD 6B
5,4,4	CARD 6C
S	CARD 6D
2	CARD 7
'A',41,42,43	CARD 7.1
'A',44,45,46	CARD 7.1
2	CARD 8
4265.5.,4266.	CARD 8.1
4635.,4645.	CARD 8.1
2.,0.	CARD 9
N	CARD 10
3,30	CARD 10.1
0	CARD 11
14,13	CARD 12
14,13	CARD 12
14,13	CARD 12
14,13	CARD 12
2	CARD 13
2	CARD 13.1
3	CARD 13.1
2	CARD 14
'E'	CARD 14.1
15,16,19	CARD 14.2
8	CARD 14.3
sb3G21X2.DER	CARD 14.4
1,1.,1	CARD 14.5
sb3G21X4.DER	CARD 14.4
1,1.,1	CARD 14.5
sb2G21X2.DER	CARD 14.4
1,1.,2	CARD 14.5
sb2G21X4.DER	CARD 14.4
1,1.,2	CARD 14.5
sb1G21X2.DER	CARD 14.4
1,1.,3	CARD 14.5
sb1G21X4.DER	CARD 14.4
1,1.,3	CARD 14.5
snkG21X2.DER	CARD 14.4
1,1.,4	CARD 14.5
snkG21X4.DER	CARD 14.4
1,1.,4	CARD 14.5
'E'	CARD 14.1
17,18,20	CARD 14.2
4	CARD 14.3



sb3G22X2.DER	CARD 14.4
2,1.,1	CARD 14.5
sb2G22X2.DER	CARD 14.4
2,1.,2	CARD 14.5
sb1G22X2.DER	CARD 14.4
2,1.,3	CARD 14.5
snkG22X2.DER	CARD 14.4
2,1.,4	CARD 14.5
3	CARD 15
'G'	CARD 15.1
15,17,31,16,37,37,18,37,32	CARD 15.2
8	CARD 15.3
sb3G31P2L.DER	CARD 15.4
1,1.,1	CARD 15.5
sb3G31TGA.DER	CARD 15.4
1,1.,1	CARD 15.5
sb2G31P2L.DER	CARD 15.4
1,1.,2	CARD 15.5
sb2G31TGA.DER	CARD 15.4
1,1.,2	CARD 15.5
sb1G31P2L.DER	CARD 15.4
1,1.,3	CARD 15.5
sb1G31TGA.DER	CARD 15.4
1,1.,3	CARD 15.5
snkG31P2L.DER	CARD 15.4
1,1.,4	CARD 15.5
snkG31TGA.DER	CARD 15.4
1,1.,4	CARD 15.5
'G'	CARD 15.1
15,15,33,16,37,37,16,37,34	CARD 15.2
4	CARD 15.3
sb3G32P2L.DER	CARD 15.4
1,2.,1	CARD 15.5
sb2G31P2L.DER	CARD 15.4
1,2.,2	CARD 15.5
sb1G31P2L.DER	CARD 15.4
1,2.,3	CARD 15.5
snkG31P2L.DER	CARD 15.4
1,2.,4	CARD 15.5
'G'	CARD 15.1
15,17,35,16,37,37,18,37,36	CARD 15.2
8	CARD 15.3
sb3G33P2L.DER	CARD 15.4
1,1.,1	CARD 15.5
sb3G33TGA.DER	CARD 15.4
1,1.,1	CARD 15.5
sb2G33P2L.DER	CARD 15.4
1,1.,2	CARD 15.5
sb2G33TGA.DER	CARD 15.4
1,1.,2	CARD 15.5
sb1G33P2L.DER	CARD 15.4
1,1.,3	CARD 15.5
sb1G33TGA.DER	CARD 15.4
1,1.,3	CARD 15.5
snkG33P2L.DER	CARD 15.4
1,1.,4	CARD 15.5
snkG33TGA.DER	CARD 15.4
1,1.,4	CARD 15.5
0	CARD 16
6	CARD 17
7	CARD 17
8	CARD 17

9		CARD 17
51		CARD 18
52		CARD 18
53		CARD 18
54		CARD 18
Y		CARD 19
1.0,16.0,1.		CARD 19.1
7		CARD 20
4		CARD 20
0.		CARD 21
0.00022		CARD 21.1
Y		CARD 22
Y		CARD 23
N		CARD 24
SET ERRdef	5.E-10	MINUIT command
SIMPLEX	3000	-
SET ERRdef	1.E-10	MINUIT command
SIMPLEX	6000	-
EXIT		-

## 8. Estimation of statistical errors

### 8.1. Introduction

The statistical theory provides rigorous prescriptions for the determination of errors associated with the random noise in the data in the framework of a non-linear minimization scheme [22, 16, 14]. The application of these procedures to the XAFS data analysis is also described in details elsewhere [11, 12].

In brief, a residual fitting function is defined as:

$$R(\lambda) = \sum_i \frac{[\alpha(E_i) - \alpha_m(E_i, \lambda)]^2}{\sigma_i^2}$$

where  $\alpha_m(E_i, \lambda)$  is the model spectrum depending on a set of fitting parameters  $\lambda_j$ ;  $j = 1, \dots, p$ . The sum is extended over all unbiased experimental points  $i$ .  $N$  is the total number of experimental points and  $p$  the number of fitting parameters. In `fittheo` the inverse squared noise  $1/\sigma_i^2$  is modeled as a smooth function of the wave-vector  $k$  like  $k^w$  (with a suitable choice of the real power  $w$ ) and the residual is calculated as

$$R'(\lambda) = \frac{\sum_i [\alpha(E_i) - \alpha_m(E_i, \lambda)]^2 k_i^w}{\sum_i \alpha_N(E_i)^2 k_i^w}$$

representing a sort of squared average signal to noise ratio. The scaling factor between  $R'(\lambda)$  and  $R(\lambda)$  is taken into account in the successive evaluation.

The correct choice of the weight  $w$  can be checked looking at the `fittheo` output plot `PS'ABC'.DAT`. The program calculates the average noise variance  $\sigma^2$  over successive regions of the spectrum, as shown in the first plot. If the trend of these data is followed by the inverse weight function corresponding to  $w$  (shown as a dashed line) the  $w$  choice can be assumed to be correct. The correctness of the random noise estimation can be checked looking at the accuracy of the interpolation function of the spectrum. For poorly sampled spectra the algorithm may fail and the noise can result underestimated (or overestimated). In this case the actual numbers suggested for the statistical tests should be taken with caution or appropriately scaled.

The statistical analysis is based on the following statements:

- 1) The optimal structural parameters are the set of  $p$  numbers  $\lambda$  corresponding to the minimum of the residual function  $R(\lambda)$ .
- 2) If the structural model is correct and the signal has been appropriately calculated, then the residual function is a  $\chi^2$  distribution (in the space of the possible realizations of the experimental noise) with  $N-p$  degrees of freedom that is:  $R(\lambda) = \chi_{N-p}^2$ .
- 3) The confidence interval in the combined  $p$ -parameter space is defined as the region enclosed by the  $(p-1)$ -dimensional surface defined by the equation  $R(\lambda) = R_{\min} + C$ .  $C$  is a critical value for the  $\chi_p^2$  random variable with a specified confidence level (usually 95%).

The set of relevant parameters involved in the determination of the joint confidence interval depends on the actual scientific problem. If for instance the focus is on a single parameter irrespective of the value of the others  $p=1$ . In more realistic cases the fit will involve the joint optimization of structural and empirical parameters all of which are relevant to the scientific question and should be included in the  $p$  count. On the other hand, the fit will usually also involve a few parameters defining the background shape. These parameters are, usually, both irrelevant and largely uncorrelated with the structural ones, so they can be safely fixed before the final minimization procedure and omitted from the  $p$  count. This procedure is equivalent to perform a preliminary extraction of the XAFS signal.

The shape of the confidence intervals in the  $p$  parameter space can be very complex and difficult to visualize. In the limit of a parabolic minimum the shape is however a multi dimensional ellipsoid whose shape is fully defined by the Hessian matrix of the residual function in the minimum. The shape of the confidence interval region can also be inferred looking at the intersection with any two-parameter plane. The values of  $R(\lambda)$  as a function of any two parameters can be visualized as a two-dimensional contour plot. The curves of equal  $R$  will be roughly elliptical. For a correct evaluation of the statistical errors, it is recommended always to visualize contour plots between highly correlated parameters such as for example  $S_0^2$ ,  $\sigma^2$  or  $E_0$ ,  $R$ . Several examples of contour plots can be found in previous works (see GNXAS publication list in Appendix A2).

## 8.2. Practical instructions for error analysis

The program `fittheo` performs the non-linear fitting minimization using the MINUIT subroutine of the CERN library [14]. At the end of the input file a series of MINUIT commands can be used to select minimization algorithms, fix and release parameters, calculate the Hessian matrix and also view two-dimensional contour plots. This should be done using the following guidelines.

- 1) There is no sense to try to perform an error analysis unless a good minimum has been obtained and without any confidence that the experimental signal has been interpreted using a correct structural model. All the experimental structural signal in the frequency range of the included signals should be reproduced by the model spectrum. If this is not the case the fit has to be improved first. Typical commands to achieve a preliminary minimum are

```
SET ERRDEF          1.E-10
SIMPLEX
```

The ERRDEF value defines convergence criteria for the routine typical values are in the range 1.E-8, 1.E-11.

- 2) The spline coefficients, that are normally refined simultaneously to MINUIT parameters should be fixed to the optimal values before the final minimization procedure is performed. This is done with the special `fitheo` command

```
CALL FCN      5
```

that can be inserted after a preliminary minimization procedure. Due to the limited numerical accuracy, which is a machine dependent problem, the fixed spline may occasionally deviate from the optimal value. A small change in the input parameter values or minimization procedure should be sufficient to solve the problem.

- 3) The final minimization should be performed only on the relevant structural parameters and, possibly,  $E_0$  and  $S_0^2$ , it is therefore recommended that all the background parameters associated with possible double-electron edges are fixed. This is done using the MINUIT command

```
FIX          NP
```

several times for each corresponding parameter number NP.

- 4) The final minimization procedure involving all the  $p$  relevant parameters is aimed at finding both minimum and covariance matrix. This is performed by the MINUIT commands

```
SET ERRDEF      5.E-6
MIGRAD
```

This minimization routine hardly converges if the spline is still being refined, and, for this reason, the spline has to be fixed according to the previous point (2). MIGRAD has more severe convergence criteria than SIMPLEX, for this reason the ERRDEF level has to be reset to a value usually about a few orders of magnitude higher. Typical values in the range 1.E-3, 1.E-6 have been found useful. In the optimal conditions MIGRAD converges and produces a covariance matrix. If the matrix is inaccurate it can be recalculated with the command

```
HESSE
```

Notice that in a large-dimensionality parameter space, correct covariance matrices are hardly generated, also due, probably, to numerical instabilities. A reasonable correlation matrix has off-diagonal numbers in the range 0.1–0.9. The matrix can not be considered correct when it contains many 0.000, or values with modulus very near to 1.000. The covariance matrix may happen to be not positive defined, this is usually a sign of being at a large distance from the actual minimum or may indicate a bad choice of fitting parameters (for instance fitting  $S_0^2$  and coordination numbers simultaneously will certainly give such a condition). The preliminary minimization should be improved and revisited. If the troubles in the covariance matrix originate from numerical inaccuracies attempts can be done changing the ERRDEF value and iterating the MIGRAD command until a good minimum is found.

- 5) In order that the printed covariance matrix represents the errors the ERRDEF value has to be set to the value corresponding to the critical value ( $C$ ) of the  $\chi_p^2$  random variable, according to the prescriptions of the point (3) in the previous section. The suggested ERRDEF value for error determination is reported in the \*.out `fitheo`

output file and can be read after a first minimization run. The value and the correct covariance error matrix are then generated adding the commands.

```
SET ERRDEF      error
HESSE
```

Here `error` represents the critical value  $C$  of the  $\chi_p^2$  random variable. The correct choice of the `ERRDEF` value `error` is crucial and the suggested value is realistic only if the estimation of the noise in the data (plot `ps`) and the  $w$  choice are correct. If the noise estimation fails but the fit is excellent it is reasonable to assume that the residual in the minimum is given by experimental noise. Then the `ERRDEF` value can be estimated by scaling the suggested `ERRDEF` value by the ratio between the actual residual in the minimum and its expected value also reported in the `*.out` file.

- 6) The contour plots representing the value of the  $R(\lambda)$  function in any 2-dimensional parameter subspace can be generated with the command:

```
CONTOUR      N1      N2
```

The contour plots involving parameters `N1` and `N2` will be printed as 25x25 character maps in the output file. The region enclosed by the “1” symbols will be the desired intersection between the  $p$ -dimensional 95% confidence interval and the chosen two-parameter sub-space.

- 7) After checking that the desired contours have the appropriate scales and are meaningful, it is possible to convert the character plot into an improved TOPdrawer output [21] using the program `contour`. The usage is straightforward. First the output has to be saved into a file then `contour` (requiring two input lines) can be executed, in UNIX

```
fitheo < fit.in > fit.out
contour
fit.out
abc
```

In VMS supposing that the log file of the `FITHEO` run is saved into `MYFILE.LOG`.

```
$ R CONTOUR
MYFILE.LOG
ABC
$ EXIT
```

The strings `abc` or `ABC` are used to generate the names of the output TOPdrawer files. The program identifies the contour plots maps and interpolates the curves producing a set of output files named `abc.ct1`, `abc.ct2`,... up to a maximum of 9 plots. Notice that since the resolution and accuracy is that of the original standard 25x25 character maps, the TOPdrawer plots will still be affected by this resolution limits.

The `*.ctn` plots might contain spikes and curves inappropriately joined, this is because of the complexity of the algorithm. They can obviously be edited and corrected if an improved presentation is required. Please check always the consistency between `*.ctn` plot and original map in order to identify the appropriate contour and also to check the correctness of the  $x$  and  $y$  scales.

- 8) For small  $p$  the estimate of the statistical errors can be performed determining the extension of the confidence intervals in the contour plots. Notice that if a single parameter is highly correlated with many others, then the real maximum extension of the confidence interval in that direction is much larger than the single extension in each

contour plot. It is advised that both information on the extension of the confidence interval and on the parameter correlation is given in the report. The indication of the maximum extension of the intervals is an incomplete information because the real  $p$ -dimensional confidence intervals are not defined by planes perpendicular to the parameter axis. Reporting as the error only the intersection of the confidence interval with the axis of each parameters is a wrong procedure because no account for parameter correlation is taken, and the errors result highly underestimated.

## 9. XAS data-analysis of disordered systems

### 9.1. Introduction

In homogeneous condensed systems the average distribution of atoms can be expressed by a collection of partial radial distribution functions  $g_{\alpha\beta}(r)$ . The average number of atoms of type  $\beta$  surrounding atoms of type  $\alpha$  in a shell of width  $dr$  at a distance  $r$  is given by

$$dN = 4\pi r^2 \rho_\beta g_{\alpha\beta}(r) dr$$

if the positions of the atoms are uncorrelated, as it is the case for systems without long range order at long distance,  $g_{\alpha\beta}(r) \rightarrow 1$  and their radial density coincides with the macroscopic density  $\rho_\beta$ .

In the case of crystalline systems or for molecules the distributions can be modeled by the sum of independent isolated shells, each one parameterized by a suitable functional form dependent on a set of parameters representing position, width, intensity, and shape.

The validity of such a model for the local neighborhood allows us to fit the XAFS with a sum of oscillating contributions each associated with a particular shell as it is performed by the program `fittheo`. The calculated signals are basically those associated with a singular distribution of atoms with small amplitude damping and phase correction associated with the finite size. For disordered systems the width of the distribution broadens and usually the peaks overlap leading to a radial distribution function always positive above a closest approach distance. If  $\gamma_{\alpha\beta}(k, r)$  is the XAFS signal produced by an atom at distance  $r$ , the total XAFS pair-signal is given by:

$$\chi(k) = \sum_{\beta} \int_0^{\infty} 4\pi r^2 \rho_{\beta} \gamma_{\alpha\beta}(k, r) g_{\alpha\beta}(r) dr$$

As pointed out by Babanov and coworkers [23, 24], it is not possible to use the previous expression for fitting the radial distribution functions on the XAFS. The inverse problem is largely unconditioned and ill-posed.

In the framework of the GNXAS data-analysis method we have proposed a simple and rigorous way of treating the previous expression in the framework of an effective peak fitting procedure (see ref. [20]). This procedure has been described in details for a single component system and can be summarized by the following steps:

- 1) The starting point is a model  $g(r)$  decomposed into one or more short distance peaks plus a long range tail  $g(r) = g_1(r) + g_2(r) + g_t(r)$ . This decomposition can be done with the help of the program `grfit`
- 2) The signal associated with  $g_i(r)$  is calculated by integration using the program `grxas`. This signal is stored in a \*.DER file like those associated with typical distances for the short range peaks calculated using `gnxas`.

- 3) A peak fitting is performed using `fittheo` refining the parameters of the short range peaks while keeping constant the tail contribution. The idea is that, being XAFS not sensitive to the long range order, the information contained in the model should not be altered. The fitting is moreover performed with constraints in the total coordination number and second moment of the matter distribution. This implies keeping constant the limit  $k \rightarrow 0$  of the  $S(k)$  function and its curvature, corresponding to long range sum-rules which can hardly be improved with the XAFS information.
- 4) Finally the fitted  $g(r)$  is reconstructed with the help of the program `grrec`.

An example of application of the method to multi-component disordered systems, such as molten salts, is contained in ref. [25]). The usage of the above mentioned programs in the case of decompositions with one or two short range peaks is described in details below.

## 9.2. Program `grfit`

### 9.2.1. Input files/parameters

This program requires an input radial distribution file. This file should contain two columns  $r$ ,  $g(r)$  for a suitable  $r$  mesh and is written in free format. A typical  $r$  mesh will have a spacing in the range  $\Delta r = 0.01-0.1 \text{ \AA}$  on the first peak which may increase on the tail up to  $\Delta r = 0.1-0.2 \text{ \AA}$ . The  $g(r)$  should be tabulated up to about 8–12  $\text{\AA}$ .

The program runs with the command

```
grfit < grfit.in
```

under UNIX or using a suitable command file under OpenVMS (see also Appendix A3). The file `grfit.in` is an input for the MINUIT subroutine of the CERN library with extra specific inputs. An example of such an input file for the  $g(r)$  of liquid Pd contained in the file `pd_was.gdr`, as tabulated from Waseda, is reported below.

```

** G(R) PEAKS FITTING PARAMETERS **
  1  SO_1    9.731    0.2    0.000    15.10
  2  Si_1    .0903   .10    0.01    0.7
  3  Er_1    2.803    .1    2.40    4.2
  4  b_1     0.829   .010   .00    2.0
  5  SO_2    3.809    0.5    0.000   9.0
  6  Si_2    .1808   .10    0.01    0.8
  7  Er_2    3.810   .10    3.2    4.8
  8  b_2     0.87    .010   .30    1.5

pd_was.gdr
0.0594          ! rho atoms/A**3
2              ! # of peaks
'G',1,2,3,4    ! type and number of parameters for peak1
'G',5,6,7,8    ! type and number of parameters for peak2
0.,3.42        ! FIT interval
0.,10.         ! PLOT interval
PdL
SET ERRDEF     .000000001
SIMPLEX        4000
SET ERRDEF     .00001
MIGRAD
EXIT
```

The explanation of the various lines is as follows:

- 1–9 the first 8 lines after the comment line are the MINUIT parameters, each line reports number, name, initial value, initial step, minimum and maximum value for the minimization range. WARNING: These lines are read with a format, each field is 10 character long. Make sure you have not typed a “tab” character!!!
- 10 A blank line indicates the end of the parameter listing.
- 11 Input  $g(r)$  file name.
- 12 Corresponding density in atoms/Å<sup>3</sup>.
- 13 Number of short range peaks to be fitted (1,2,3,...).
- 14–15 For each peak, report peak type 'G' and number of previously defined MINUIT parameters corresponding to intensity (Coordination Number), Variance  $\sigma^2$ , average distance  $R$  and skewness (asymmetry parameter)  $\beta$ . The default profile 'G' is that of a Gamma distribution in the  $4\pi r^2 \rho g(r)$  function. The Gaussian limit is obtained in the limit  $\beta=0$ . Note that this will not correspond to a Gaussian shape in the  $g(r)$  unless  $\sigma^2 \ll R^2$ .
- 16 Fitting  $r$  range (Å).
- 17 Plot  $r$  range (Å).
- 18 (A3) string used to generate automatically the output file names.
- 19–23 MINUIT commands: at least one minimization command should be given (*i.e.* SIMPLEX), the last command should be EXIT.

### 9.2.2. Output files/parameters

The output of the program execution, written on the standard output, contains usual information of the MINUIT minimization and some additional information of the input and output results. The main output results are contained in 4 files whose name is composed by the (A3) string name followed by different extensions.

- \*.GRF Is the main fitting plot (for TOPdrawer), it contains the original  $g(r)$  data and the profiles of the fitted peak(s) and that of the remaining tail.
- \*.TAI This file contains the  $g(r)$  tail data printed in two columns free format. These data correspond to the original data minus the fitted peak profile. All points after the first negative value found going in the direction of decreasing  $r$  are set to zero. (Noise and fitting errors are not evident, please be careful). This file is suitable for further processing with the program `grxas`.
- \*.SDK This file suitable for TOPdrawer is a plot of the structure factor  $S(k)$  corresponding to the input  $g(r)$  and that of the fitted model. This latter plot for isolated peaks is meaningless.
- \*.OUT This is a text file containing the most relevant output numerical information (which is also printed on the standard output at the end of the program run). In particular, for each fitted peak the values of the optimal parameters for  $N$ ,  $R$ ,  $\sigma^2$ , and  $\beta$  are reported. Then, for the user convenience the values to impose the sum-rule constraints on the total coordination number and total second moment are reported. For the case of two short range peaks these values correspond to

$$\begin{cases} N = N_1 + N_2 \\ M^{(2)} = N_1(R_1^2 + \sigma_1^2) + N_2(R_2^2 + \sigma_2^2) \end{cases} \quad (1)$$

=====



```

Number of short range peaks = 2
      N      R      sigma**2      beta
    9.6273   2.79925   0.08468   0.8141
    2.7121   3.46024   0.06677   0.6579
=====
SUM RULE CONSTRAINTS FOR g(r)
NUMBER OF ISOLATED PEAKS = 2
TOTAL COORDINATION NUMBER = 12.3393526
TOTAL SECOND MOMENT      = 108.906441
=====

```

### 9.2.3. The gamma function profile

This functional form was found to be extremely versatile and able to model atomic distributions in a large number of cases. We have fitted the atomic distributions resulting from Molecular Dynamics or Monte Carlo simulations [26–27] of solids in a wide temperature range using a gamma-function profile for each coordination shell. The success of this model resides, in this case, in the possibility to tune the asymmetry from the Gaussian limit to a moderate asymmetry limit. The application to the EXAFS analysis is further simplified by the fact that the configurational average over such a model distribution results in a simple analytical amplitude and phase correction. Moreover the distribution has a correct probabilistic interpretation and does not fall in the drawbacks of alternative approaches like the cumulant expansion method. In the case of liquid systems, the gamma functional form is found to be a versatile model peak to be included in a short range model  $g(r)$  [18–19].

The gamma distribution model is an asymmetric peak for the  $4\pi\rho r^2g(r)$  function defined by the four parameters:

$N_c$  – coordination number,

$R$  – average distance,

$\sigma$  – standard deviation,

$\beta$  – skewness, (third moment about the mean divided by  $\sigma^3$ ).

The gamma distribution can be written analytically as

$$f(r) = \frac{2 N_c}{\sigma |\beta| \Gamma(4\beta^{-2})} \left[ 4\beta^{-2} + \frac{2(r-R)}{\beta\sigma} \right]^{(4\beta^{-2}-1)} \exp \left[ -\left( 4\beta^{-2} + \frac{2(r-R)}{\beta\sigma} \right) \right] \quad (2)$$

where  $\Gamma(x)$  is the Euler's Gamma function. The previous expression is valid for  $(r-R)\beta \geq -2\sigma$  and  $f(r) = 0$  elsewhere. This function approaches the Gaussian limit when  $\beta \rightarrow 0$ . It is however defined in a wide interval of positive and negative asymmetry, in particular the probability density does not diverge in the range  $-2 \leq \beta \leq 2$  and is continuous in the range  $|\beta| < 2$ . The best range for physical applications is however  $|\beta| \leq \sqrt{2}$  where the first derivative is also continuous for every  $r$ .

## 9.3. Program grxas

### 9.3.1. input files/parameters

Three input files are required to run the program `grxas`. These are: the phase-shifts (`.t1`) and atomic absorption (`.s0`) files as calculated by the program `phagen` and any two columns free format ASCII file containing the values  $r$ ,  $g(r)$  for a given radial distribution model or tail. The algorithm used is that of summing the single contributions for each  $r$  point with an appropriate weight and a small Gaussian damping (for smoothing) depending on the

$\Delta r$  spacing. In practice a smoothed trapezoidal integration algorithm is used which should be compatible with variable  $r$  spacing and realistic tabulated  $g(r)$  values. The program runs with the command

```
grxas < grxas.in
```

under UNIX or using a suitable command file under OpenVMS (see also Appendix A3). The file `grxas.in` contains all input parameters for the calculation. An example of such an input file to calculate the signal associated with the tail of the liquid Pd  $g(r)$  as decomposed in the previous `grfit` example is reported below.

```

../pha/pdxpha.tl
../pha/pdxpha.s0
ptw          ! conventional name output files
2            ! (1) TOTAL g2 CONTRIBUTION, (2) X2 TERM ONLY
10.         ! > DISTANCE CUTOFF (A) in X calculations
0           ! > DERIVATIVE CALCULATION? (1)
.0594,8.,.4 ! > sample density atoms/A**3, cutoff e standard deviation A
0.0,0.8,0.9 ! > Gamma hole (HWHM eV), resolution (sigma eV) and S02
1,2,1,2     ! > PHASE1, PHASE2, LABEL1, LABEL2
../grf/PdL.TAI
0           ! > Core state angular momentum, LIN
.0001,6,.000001,80 ! > toll_lmax, Mmax, EPS, NEP
1          ! > OUTPUT FLAG: IOUT
0.         ! > K weight in plots
0.1,15.,.05 ! > K min, K Max, Delta k interpolation
.4         ! shift for energy scale
2          ! R choice in .DER: (1) first maximum of r2g(r)

```

The description of the input lines (very similar to those for `gnxas`) is as follows:

- 1–2 Names of the phase-shifts and atomic cross-section files as calculated by `phagen`.
- 3 (A3) string used to generate automatically the output file names.
- 4 flag to choose either the continued fraction algorithm (1) to calculate  $\gamma^{(2)}$  signals or (2) the single scattering approximation  $\chi_2$  (which is perfectly adequate in most of the cases).
- 5 Maximum distance cutoff for signal calculation.
- 6 flag to activate the calculation of the derivative with respect to a distance scaling parameter. Put 0 as a default if you wish only to calculate the signal associated with a given  $g(r)$ .
- 7 Density (atoms/Å<sup>3</sup>) of the corresponding  $g(r)$ , distance  $R_c$ , and standard deviation  $\sigma_c$  of the smooth cutoff. In practice the  $g(r)$  integration cannot be sharply truncated. The upper integration limit must be smoothed out. This is done by multiplying the  $r^2g(r)$  function by a semi-Gaussian function for  $r > R_c$  with standard deviation  $\sigma_c$ . Provided that  $R_c$  and  $\sigma_c$  are large enough the signal will not be affected. The signal calculation cutoff (line 5) must be at least  $\geq R_c + 3\sigma_c$ .
- 8 Physical and experimental damping of the signal: Additional core-hole half-width half-maximum Lorentzian width (put 0. if already included in the complex phase-shifts), Gaussian experimental energy resolution width (standard deviation eV), amplitude reduction factor. WARNING: All these parameters have to coincide with those used in the fitting if one wishes to reproduce exactly the fixed tail signal contribution.
- 9 Phaseshift indexes corresponding to photoabsorber and backscattering atoms and conventional labels.

- 10  $g(r)$  file name. In this case it is the liquid Pd tail.
- 11 Value for the initial state angular momentum (0 for  $K, L_I$ -edges, 1 for  $L_{II}, L_{III}$ -edges)
- 12 Convergence parameters and number of energy points of the phase shift range to be used in the calculation.
- 13 Output flag (keep 1)
- 14–16  $k$  weight,  $k$  range and energy shift (to make the theoretical energy values always positive) of interpolated output plots. These numbers do not affect the calculation.
- 17 Very important calculation flag. The calculated signal will be stored in a \*.DER file similarly to those generated by `gnxas`. These files contain the information of a reference distance for which the signal has been calculated. In the  $g(r)$  case this is meaningless, but in order that `fitheo` generates the signal without any alteration, the effective distance MINUIT parameter used in the fit must be fixed to the number written in the \*.DER file by `grxas`. Two possibilities can be chosen (1) write in the \*.DER file the position of the first maximum of the  $r^2g(r)$  function (2) write in the \*.DER file the distance corresponding to the leading frequency of the resulting  $\chi(k)$  oscillations. WARNING: a check that the selected algorithm has not failed is recommended.

### 9.3.2. Output files/parameters

The output contains information on the calculation for each distance point. The very important information on the distance number adopted and written in the \*.DER file is written at the end of the execution. Two output files are produced. Let be XXX the (A3) string in the input command file, then the output files will be:

XXXGR.DER This is the \*.DER file to be used as input in the fitting.

XXXGR.PLO This is a TOPdrawer plot containing the calculated signal.

### 9.4. Specific information for $g(r)$ refinements using `fitheo`

This information integrates what is written in the `fitheo` program documentation. The “peak-fitting” approach used by `fitheo` can be successfully used in liquid systems if appropriate constraints accounting for the long-range behavior are introduced. The way of FIXing the coordination number (CN) or the sum of two coordination numbers is already contained in the `fitheo` documentation (see explanation of cards 13 and 13.1). The tail contribution and the constrained peak refinement using the second-moment constraint

$$\Delta [N_1(R_1^2 + \sigma_1^2) + N_2(R_2^2 + \sigma_2^2)] = 0$$

which can be applied to the refinement of two asymmetric peaks with parameters  $N_1$ ,  $R_1$ ,  $\sigma_1^2$ ,  $\beta_1$ ,  $N_2$ ,  $R_2$ ,  $\sigma_2^2$ , and  $\beta_2$  is accounted for by `fitheo` writing simple input parameters in the `fitheo` input command file.

Let us suppose to have a model  $g(r)$  decomposed into two  $\Gamma$ -like or Gaussian functions plus a long-range tail. We can then calculate the constraints  $N = N_1 + N_2$  and  $M_2 = [N_1(R_1^2 + \sigma_1^2) + N_2(R_2^2 + \sigma_2^2)]$  using for example the program `grfit`. Those constraints must be assigned as MINUIT parameters in the list and kept FIXed during the minimization.

In order to refine a model  $g(r)$  composed of two short range peaks and a tail several specific MINUIT parameters are required. Parameter numbers 15-16-17, 25-26-27, and 30-31

in the example below refer to these three contributions, respectively. Parameters 50-52-54 refer to coordination numbers.

15	R1	2.83280	.015	2.800	2.870
16	S1**2	.080022	.0010	0.060000	0.1000
17	beta1	.8687	.01	0.500	1.0
25	R2	3.34550	.015	3.300	3.800
26	S2**2	.095160	.0010	0.020000	0.4000
27	beta2	0.9214	.01	0.500	1.2
30	Rt	3.35501	.000	2.840	5.256
31	St**2	.000000	.0000	0.000000	0.0900
50	N	12.33935	0.00	0.90	19.7
52	Nc3	1.000	0.00	0.90	2.1
54	M2	108.9064	0.00	0.00	250.00

Only the first 6 parameters are free and they define the shape of the first two peaks. The corresponding coordination numbers are calculated by applying the constraints on the total coordination number and second moment (parameters 50 and 54 in the present example), by solving the system (Equation 1) of two linear equations with two unknowns  $N_1$  and  $N_2$ . In order to tell the program that a special coordination number definition is required the following input cards are included in the coordination number definition section of the fitheo input:

```

3 ! CARD 13 number of coordination numbers
50 ! CARD 13.1 parameter corresponding to the first number
-1 ! CARD 13.1 flag for second moment constraint
54,15,16,25,26 ! CARD 13.1.1 parameters for constraints M2,R1,S1**2,R2,S2**2
52 ! CARD 13.1 parameter corresponding to the third coordination number

```

The second parameter number  $-1$  activates a hidden flag in the program, under these conditions the parameter number in the previous line (50 in this example) is considered as  $N_1 + N_2$ , in addition another input line (card 13.1.1) with 5 parameter numbers is required. These parameters must be the one for the fixed value of the second moment  $R_1$ ,  $\sigma_1^2$ ,  $R_2$ , and  $\sigma_2^2$ . With the flag activated, the actual first and second coordination numbers will be calculated at each step as the solution of the system (Equation 1) of two linear equations with two unknowns. The final line with parameter number 52 corresponds to a fixed parameter with coordination number 1.0. This will be the effective coordination number used in the calculation (grxas) of the tail signal. Notice that the tail signal is also averaged over a Gaussian distribution with  $\sigma^2 = 0$ . The way all these parameters are used for averaging the three signal contributions is clarified by the following fitheo input parameters (cards 14, 14.1-5):

```

3 ! number of g2 peaks
'E'
15,16,17 ! R, sigma**2, beta for 1st shell
1 ! number of signals
../xas/pd1G21X2.DER
1,1. ! coordination number, and additional degeneracy
'E'
25,26,27
1 ! number of signals
../xas/pd1G22X2.DER
2,1. ! coordination number, and additional degeneracy
'G'
30,31

```

```
1
  ../../xas/ptwGR.DER
3,1.
```

Among the fitting output there will be a file named `grrXXX.out` containing the following lines summarizing the fitting output parameters for the three peaks. Some of these lines are ready to be included in the input of the program `grrec` that will be used to draw the complete optimized  $g(r)$  function.

```
3
'E'
  8.8364325      2.8008789  0.6880615E-01    0.8033783
'E'
  3.5029171      3.3218813  0.9219276E-01    0.8274093
'G'
  1.0000000      3.3550100  0.0000000E+00    0.0000000
```

## 9.5. Program `grrec`

### 9.5.1. Input files/parameters

This program is suitable to compare tabulated and refined  $g(r)$  models. The input files are two columns ASCII files reporting the values  $r, g(r)$  either for a complete  $g(r)$  or for a tail. The standard input commands required to run the program to compare a tabulated  $g(r)$  with a reconstructed  $g(r)$  for a liquid system and with that for an f.c.c. solid (first three shells), is as follows.

```
../../grf/pd_was.gdr
psd
2.,3.4
2
.0594
../../grf/PdL.TAI
2
'E'
  8.8364325      2.8008789  0.6880615E-01    0.8033783
'E'
  3.5029171      3.3218813  0.9219276E-01    0.8274093
.0636
xxxxxx
3
'E'
  12.0000000     2.8165597  0.5145272E-01    0.5966927
'G'
  6.0000000      3.8337270  0.4290504E-01    0.0000000
'G'
  24.0000000     4.7310289  0.6966901E-01    0.0000000
```

- 1 Complete  $g(r)$  file name, this  $g(r)$  will be plotted for comparison together with the reconstructed ones.
- 2 (A3) string for the automatic generation of output file names.
- 3 Plot range (Å).
- 4 Number of models to compare (in this case 2).
- 5 Density (atoms/Å<sup>3</sup>) of the model.
- 6  $g_t(r)$  tail file name (if appropriate), if the file does not exist no tail is added to the peaks.

- 7 Number of short-range peaks.
- 8 Peak shape ('E' Gamma function, 'G' Gaussian).
- 9 Peak parameters: Coordination number, Average distance, Variance, Skewness.
- \*\* Lines 8–9 are repeated for each peak these can be directly copied from the gr-rXXX.out output file of fittheo, lines 5–\*\* are repeated for each model.

### 9.5.2. output files/parameters

This program produces a diagnostic output on the terminal and 4 output files suitable for TOPdrawer.

- \*.grf Model and reconstructed  $g(r)$  functions.
- \*.pdr Model and reconstructed  $4\pi r^2 \rho g(r)$  functions.
- \*.nдр Reconstructed running coordination number  $n(r) = 4\pi\rho \int_0^r x^2 g(x) dx$ .
- \*.sdk Model and reconstructed structure factors  $S(k)$ .

---



---

Note added:

---



---

Please consider that in the examples presented throughout the text numbers in input and output files are often reported with several decimal digits. This is only done for consistency with the writing formats and has no relation with the accuracy of the results. A proper error analysis using the MINUIT covariance matrix information should be performed following the GNXAS guidelines to establish the statistical confidence intervals in the output fitting parameters.

---



---

## References

- [1] Filipponi A, Di Cicco A and Natoli C R *Phys. Rev. B* **52**, 15122 (1995).
- [2] Filipponi A and Di Cicco A *Phys. Rev. B* **52**, 15135 (1995).
- [3] Filipponi A, Di Cicco A, Tyson T A and C. R. Natoli, *Solid State Commun.* **78**, 265 (1991).
- [4] Johnson K H *Adv. Quantum Chem.* **7**, 143 (1973).
- [5] Natoli C R and Benfatto M *J. Phys.* (Paris), C8, 11 (1986).
- [6] Tyson T A, Hodgson K O, Natoli C R and Benfatto M *Phys. Rev. B* **46**, 5997 (1992).
- [7] S.-H. Chou, J. J. Rehr, E. A. Stern, and E. R. Davidson, *Phys. Rev. B* **35**, 2604 (1987).
- [8] Mustre de Leon J, Rehr J J, Zabinsky S I and Albers R C *Phys. Rev. B* **113**, 5135 (1991).
- [9] Desclaux J P *Comp. Phys. Commun.* **9**, 31 (1975); Desclaux J P, Mayers D F and O'Brien F J. *Phys. B* **4**, 631 (1971).
- [10] Benfatto M Natoli C R and Filipponi A *Phys. Rev. B* **40**, 9626 (1989).
- [11] Filipponi A *J. Phys. Condensed Matter* **3**, 6489 (1991).
- [12] Di Cicco A *Phys. Rev. B* **53**, 6174 (1996); Di Cicco A *J. Phys. IV* **7**, Colloque C2, 171 (1997).
- [13] MINUIT reference Manual, *Function minimization and error analysis*, CERN Program Library Long Writup D506, F. James, Version 94.1 Edition August 1998, (<http://wwwinfo.cern.ch/asdoc/minuit>).
- [14] Filipponi A *J. Phys.: Condensed Matter* **7**, 9343 (1995).
- [15] Heitler W *The Quantum Theory of Radiation*, 3rd ed. 1954, Dover publ. 1984.
- [16] Press W H, Teulosky S A, Vetterling W T and Flannery B P *Numerical Recipes in Fortran 77*, 2nd edition, Cambridge University Press, 1992, reprinted with corrections in 1999.
- [17] Filipponi A, Di Cicco A, Zanoni R, Bellatreccia M, Sessa V, Dossi C and Psaro R *Chem. Phys. Lett.* **184**, 485 (1991).
- [18] D' Angelo P, Di Nola A, Filipponi A, Pavel N V and Roccatano D *J. Chem. Phys.* **100**, (1994).
- [19] Filipponi A and Di Cicco A *Phys. Rev. B* **51**, 12322 (1995).
- [20] Filipponi A *J. Phys.: Condensed Matter* **6**, 8415 (1994).

- [21] TOPdrawer working paper of R. B. Chaffee, SLAC computation group, Stanford, CA, CGTM No. 178, August 1978 and revisions up to October 1989; J. Clement, Version 5.12, July 1993, Rice Bonner Lab, Fermilab Document PP0005.1; H. Okamura, LINUX Release 1.4c ported from Version 5.12, Department of Physics, University of Tokyo, <ftp://iris.riken.go.jp/pub/topdrawer/>.
- [22] Eadie W T, Drijard D, James F, Roos M and Sadoulet B *Statistical methods in experimental physics*, (North-Holland, 1971).
- [23] Babanov Y A, Vasin V V, Ageev A L, and Ershov N V *Phys. stat. sol. (b)* **105**, 747 (1981).
- [24] Ershov N V, Vasin V V, Ageev A L, and Babanov Y A *Phys. stat. sol. (b)* **108**, 103 (1981).
- [25] Di Cicco A, Minicucci M and Filipponi A, *Phys. Rev. Lett.* **78**, 460 (1997); *Phys. Rev. B* **56**, 11456 (1997).
- [26] Di Cicco A, Rosolen M J, Marassi R, Tossici R, Filipponi A and Rybicki J J. *Phys.: Condensed Matter* **8**, 10779 (1996).
- [27] Di Cicco A, Filipponi A, Itié J P and Polian A *Phys. Rev. B* **54**, 9086 (1996).

## Appendix

### A1. Historical developments and acknowledgments

The development of the GNXAS codes has benefited from the stimulating environment existing in the 80's at the Frascati synchrotron radiation facility where several scientists working on theoretical as well as experimental aspects were present. In particular the research effort was coordinated by C. R. Natoli (INFN, Laboratori Nazionali di Frascati), a pioneer of the multiple-scattering applications to the EXAFS spectroscopy. C. R. Natoli, together with M. Benfatto (INFN), C. Brouder (Nancy), M. F. Ruiz-Lopez (Nancy), and R. F. Pettifer (Warwick) developed the MS theory for EXAFS calculations and several pioneering codes were written by these authors.

The GNXAS code has also benefited from the fruitful collaboration with K. O. Hodgson and B. Hedman of the Stanford Synchrotron Radiation Laboratory and Stanford University, who suggested applications to the field of molecular biology and contributed to several specific developments of the codes. T. A. Tyson (formerly at Stanford) worked on the implementation of the Hedin-Lundqvist potential. P. D' Angelo and N. V. Pavel of the Department of Chemistry of the Università di Roma "La Sapienza" pioneered early applications of the codes to solution chemistry.

The main authors of the GNXAS package are A. Filipponi (Università dell' Aquila) and A. Di Cicco (Università di Camerino). C. R. Natoli and T. A. Tyson wrote the phagen code, where all the theory for the calculation of the potential is included. The `gnxas` code was mainly written by A. Filipponi, implementing efficient codes for MS calculations with fast matrix products and continued fraction expansions. The program `fitheo` has been mainly written by A. Di Cicco. Several other algorithms and approaches, implemented in the package, have been jointly developed by the main authors.

The codes are written in FORTRAN-77 and initially compiled using the VMS-FORTRAN compiler, extended successively to be compiled under the OpenVMS  $\alpha$ -VAX architecture. Since the year 1995 efforts were devoted to compile these programs under UNIX and at present a unique source code compiling under Linux, SunOS, HP-UX, OSF1 (Digital UNIX) and Open-VMS is available. Computing technical help was given by G. Mancini (Università di Camerino) for the protection software under VMS and M. F. Sanchez Del Rio (ESRF-Grenoble) for UNIX and possible IDL interface. Valuable help for the UNIX versions were recently given by N. Morgante (Milano-Bicocca) who also developed advanced Perl scripts able to handle several minimization procedures suitable for mapping the fitting residual as a function of model parameters. A preliminary version of the TeX document of the present documentation was realized by M. Minicucci (Università di Camerino).

### A2. GNXAS publication list

This is a comprehensive list of the publications related to the development and utilization of the GNXAS package of programs. It is intended to be a reference list for those interested to deepen their knowledge about characteristics and capabilities of the software described in the present documentation. There are certainly fortuitous omissions and mistakes and we apologize in advance for any missing reference or misspelled name.

#### Package

- [1] A. Filipponi, A. Di Cicco, T. A. Tyson, and C. R. Natoli, *Ab-initio modelling of X-ray Absorption spectra*, Solid State Commun. **78**, 265 (1991).
- [2] A. Filipponi, A. Di Cicco, T. A. Tyson, and C. R. Natoli, *GNXAS package documentation*, (University press, Camerino 1992).
- [3] A. Filipponi and A. Di Cicco, *Investigating local three-body correlations by means of a novel XAS data-analysis method*, Synchrotron Radiation News **6**, 13 (1993).
- [4] A. Di Cicco, *EXAFS multiple-scattering data-analysis: GNXAS methodology and applications*, Physica B **208 & 209**, 125 (1995).
- [5] A. Di Cicco, *Short-range structure in solid and liquid matter using multiple-edge EXAFS refinement*, J. Phys. IV Colloq. C2, 171 (1997).

#### Theoretical Background

- [1] M. Benfatto, C. R. Natoli, and A. Filipponi, *Thermal and structural damping of the multiple-scattering contributions to the x-ray absorption coefficient*, Phys. Rev. B **40**, 9626 (1989).



- [2] A. Filipponi, *Continued Fraction expansion for the X-ray absorption cross-section*, J. Phys.: Condensed Matter **3**, 6489 (1991).
- [3] T. A. Tyson, K. O. Hodgson, C. R. Natoli, and M. Benfatto, *General multiple-scattering scheme for the computation and interpretation of x-ray-absorption fine structure in atomic clusters with applications to SF<sub>6</sub>, GeCl<sub>4</sub>, and Br<sub>2</sub> molecules*, Phys. Rev. B **46**, 5997 (1992).
- [4] A. Filipponi, A. Di Cicco, and C. R. Natoli, *Probing Higher order correlation functions by inner shell x-ray absorption spectroscopy in disordered materials*, Proceedings of the International Conference INDIAS-91 on Disordered Materials (Structure and Properties), Ed. by S. K. Srivastava, (INDIAS Publ., Allahabad, Indore 1993) pag. 143.
- [5] A. Filipponi, A. Di Cicco, and C. R. Natoli, *X-ray absorption spectroscopy and n-body distribution functions in condensed matter (I): theory*, Phys. Rev. B **52**, 15122 (1995).
- [6] A. Filipponi and A. Di Cicco, *X-ray absorption spectroscopy and n-body distribution functions in condensed matter (II): data-analysis and applications*, Phys. Rev. B **52**, 15135 (1995).
- [7] A. Filipponi, *Radial distribution function probed by x-ray absorption spectroscopy*, J. Phys.: Condensed Matter **6**, 8415 (1994).
- [8] A. Filipponi, *Statistical errors in EXAFS data analysis*, J. Phys.: Condensed Matter **7**, 9343 (1995).
- [9] A. Di Cicco, *Multiple-edge EXAFS refinement: Short-range structure in liquid and crystalline Sn*, Phys. Rev. B **53**, 6174 (1996).

#### Applications (chronological list)

- [1] A. Filipponi, A. Di Cicco, M. Benfatto, and C. R. Natoli, *X-ray absorption spectroscopy on amorphous silicon: a probe for the three-body correlation function*, J. Non-Cryst. Solids **114**, 229 (1989).
- [2] A. Filipponi, A. Di Cicco, M. Benfatto, and C. R. Natoli, *The three-body correlation function in amorphous silicon probed by X-ray absorption spectroscopy*, Europhys. Lett. **13**, 319 (1990).
- [3] A. Filipponi, *Continued Fraction expansion for the X-ray absorption cross-section*, J. Phys.: Condensed Matter **3**, 6489 (1991).
- [4] A. Filipponi, A. Di Cicco, R. Zanoni, M. Bellatreccia, V. Sessa, C. Dossi, and R. Psaro, *Multiple scattering analysis of the x-ray absorption spectrum of Os<sub>3</sub>CO<sub>12</sub> carbonyl cluster*, Chem. Phys. Lett. **184**, 485 (1991).
- [5] A. Di Cicco, S. Stizza, A. Filipponi, F. Boscherini, and S. Mobilio, *X-ray absorption investigation of SiX<sub>4</sub> (X=Cl, F, CH<sub>3</sub>)*, J. Phys. B: At. Mol. Opt. Phys. **25**, 2309 (1992).
- [6] A. Di Cicco and A. Filipponi *Three-body distribution function in liquids: the case of liquid gallium*, J. Non-Cryst. Solids **156–158**, 102 (1993).
- [7] L. Ottaviano, A. Filipponi, A. Di Cicco, S. Santucci, and P. Picozzi, *XAS investigation of three-body correlations in liquid Hg*, J. Non-Cryst. Solids **156–158**, 112 (1993).
- [8] P. D' Angelo, A. Di Cicco, A. Filipponi, and N. V. Pavel, *Double electron excitation channels at the Br K-edge of HBr and Br<sub>2</sub>*, Phys. Rev. A **47**, 2055 (1993).
- [9] E. Burattini, P. D' Angelo, A. Di Cicco, A. Filipponi, and N. V. Pavel, *Multiple scattering x-ray absorption analysis of simple brominated hydrocarbon molecules*, J. Phys. Chem. **97**, 5486 (1993).
- [10] A. Filipponi, L. Lozzi, M. Passacantando, P. Picozzi and S. Santucci, *Three-body signature of the b.c.c. structure in extended energy-loss spectra of Cr metal*, Phys. Rev. B **47**, 8494 (1993).
- [11] G. Dalba, D. Diop, P. Fornasini, A. Kuzmin, and F. Rocca, *EXAFS and XANES study of GaAs on Ga and As K-edge*, J. Phys.: Condensed Matter **5**, 1643 (1993).
- [12] A. Filipponi, L. Ottaviano, M. Passacantando, P. Picozzi and S. Santucci, *Structural investigation of gaseous, liquid and solid Br<sub>2</sub> by x-ray absorption spectroscopy*, Phys. Rev. E **48**, 4575 (1993).
- [13] E. Nordlander, Sonny C. Lee, Wei Cen, Z. Y. Wu, C. R. Natoli, A. Di Cicco, A. Filipponi, B. Hedman, K. O. Hodgson, and R. J. Holm, *The Heterometal cuboidal clusters MFe<sub>4</sub>S<sub>6</sub>(PEt<sub>3</sub>)<sub>4</sub>Cl (M=V, Mo): Synthesis, structural analysis by crystallography and EXAFS, and relevance to the structure of iron-molybdenum cofactor of nitrogenase*, J. Am. Chem. Soc. **115**, 5549 (1993).
- [14] A. Di Cicco and M. Berrettoni, *X-Ray Absorption Multiple-scattering Study of High-Tc Superconductors*, Phys. Lett. A **176**, 375 (1993).
- [15] P. D' Angelo, A. Di Nola, A. Filipponi, N. V. Pavel, and D. Roccatano, *An EXAFS study of aqueous solutions by employing Molecular Dynamics simulations*, J. Chem. Phys. **100**, 985 (1994).

- [16] A. Filipponi, P. D' Angelo, N. V. Pavel, and A. Di Cicco, *Triplet correlations in the hydration shell of aquaions*, Chem. Phys. Lett. **225**, 150 (1994).
- [17] A. Di Cicco and A. Filipponi, *Local correlations in liquid and supercooled gallium probed by x-ray absorption spectroscopy*, Europhys. Lett. **27**, 407 (1994).
- [18] S. D. Conradson, B. K. Burgess, W. E. Newton, A. Di Cicco, A. Filipponi, Z. Y. Wu, C. R. Natoli, B. Hedman, and K. O. Hodgson, *Selenol binds to iron in nitrogenase iron-molybdenum cofactor: An extended x-ray absorption fine structure study*, Proc. Natl. Acad. Sci. USA **91**, 1290 (1994).
- [19] H. I. Liu, A. Filipponi, N. Gavini, B. K. Burgess, B. Hedman, A. Di Cicco, C. R. Natoli, and K. O. Hodgson, *EXAFS studies of FeMo-cofactor and MoFe protein - direct evidence for the long-range Mo-Fe-Fe interaction and cyanide binding to the Mo in FeMo-cofactor*, J. Am. Chem. Soc. **116**, 2418 (1994).
- [20] T. E. Westre, A. Di Cicco, A. Filipponi, C. R. Natoli, B. Hedman, E. I. Solomon, and K. O. Hodgson, *Determination of the Fe-N-O angle in FeNO<sub>7</sub> complexes using multiple scattering EXAFS analysis by GNXAS*, J. Am. Chem. Soc. **116**, 6757 (1994).
- [21] A. Di Cicco, M. Berrettoni, R. Marassi, R. Tossici, and A. Filipponi, *X-ray absorption study of molten and alkali halides*, Proceedings of the Ninth International Symposium on Molten Salts; editors C. L. Hussey, D. S. Newman, G. Mamantov, and Y. Ito, The Electrochemical Society, Proc. Vol. 94-13, p. 77 (1994).
- [22] A. Di Cicco, M. Berrettoni, S. Stizza, E. Bonetti, and G. Cocco, *Microstructural defects in nanocrystalline iron probed by x-ray absorption spectroscopy*, Phys. Rev. B **50**, 12386 (1994).
- [23] R. E. Benfield, A. Filipponi, D. T. Bowron, R. J. Newport, and S. J. Gurman, *X-ray absorption study of gold coordination compounds: EXAFS refinements and double-electron excitation background*, J. Phys.: Condensed Matter **6**, 8429 (1994).
- [24] T. E. Westre, A. Di Cicco, A. Filipponi, C. R. Natoli, B. Hedman, E. I. Solomon, and K. O. Hodgson, *GNXAS, a multiple-scattering approach to EXAFS analysis: methodology and applications to iron complexes*, J. Am. Chem. Soc. **117**, 1566 (1995).
- [25] A. Filipponi and A. Di Cicco, *Short-range order in crystalline, amorphous, liquid, and supercooled liquid germanium probed by x-ray absorption spectroscopy*, Phys. Rev. B **51**, 12322 (1995).
- [26] P. D' Angelo, A. Di Nola, E. Giglio, M. Mangoni, and N. V. Pavel, *EXAFS Study of Micellar Aggregates of Bile Acid Rubidium Salts*, J. Phys. Chem. **99**, 5471 (1995).
- [27] A. Filipponi and A. Di Cicco, *X-ray absorption spectroscopy and n-body distribution functions in condensed matter (II): data-analysis and applications*, Phys. Rev. B **52**, 15135 (1995).
- [28] T. E. Westre, A. Di Cicco, A. Filipponi, C. R. Natoli, B. Hedman, E. I. Solomon, and K. O. Hodgson, *Using GNXAS, a multiple-scattering EXAFS analysis for determination of the Fe-N-O angle in {FeNO<sub>7</sub>} complexes*, Physica B **208 & 209**, 137 (1995).
- [29] A. Di Cicco, *Probing local structure by multiple-scattering EXAFS data-analysis: application to nanocrystalline iron*, Materials Science Forum **195**, 127 (1995).
- [30] A. Di Cicco and F. Sperandini, *Cu-O-Cu in-plane angle distribution in layered superconductors probed by EXAFS*, Physica C **258**, 349 (1996).
- [31] A. Di Cicco, M. J. Rosolen, R. Marassi, R. Tossici, A. Filipponi, and J. Rybicki, *Short-range order in solid and liquid KBr probed by EXAFS*, J. Phys.: Condensed Matter **8**, 10779 (1996).
- [32] A. Di Cicco, *Multiple-edge EXAFS refinement: Short-range structure in liquid and crystalline Sn*, Phys. Rev. B **53**, 6174 (1996).
- [33] P. D' Angelo, A. Di Nola, M. Mangoni, N. V. Pavel, *An extended X-ray absorption fine structure study by employing molecular dynamics simulations: bromide ion in methanolic solution*, J. Chem. Phys. **104**, 1779 (1996).
- [34] A. Di Cicco and A. Filipponi, *Local g(r) properties in liquids probed by high-temperature EXAFS*, J. Non-Cryst. Solids **205-207**, 304 (1996).
- [35] P. D' Angelo, H.-F. Nolting, and N. V. Pavel, *Evidence for multielectron resonances at the Sr K edge*, Phys. Rev. A **53**, 798 (1996).
- [36] A. Di Cicco, A. Filipponi, J. P. Itié, and A. Polian, *High pressure EXAFS measurements of solid and liquid Kr*, Phys. Rev. B **54**, 9086 (1996).
- [37] M. Giorgetti, M. Berrettoni, P. Conti, A. Di Cicco, I. Ascone, *X-ray absorption spectroscopic study of "Costa-Type" Organocobalt Coenzyme B<sub>12</sub> models*, Organometallics **15**, 3491 (1996).

- [38] H. H. Zhang, A. Filipponi, A. Di Cicco, S. C. Lee, M. J. Scott, R. H. Holm, B. Hedman, and K. O. Hodgson, *Multiple-edge XAS Studies of synthetic Iron-Copper bridged molecular assemblies relevant to Cytochrome c Oxidase. Structure determination using multiple-scattering analysis with statistical error evaluation of errors*, Inorg. Chem. **36**, 4819 (1996).
- [39] A. Filipponi, *Short-range order in liquid matter probed by high-temperature x-ray absorption measurements*, J. Phys.: Condensed Matter **8**, 9335 (1996).
- [40] A. Di Cicco, *Local Structure in binary liquids probed by EXAFS*, J. Phys.: Condens. Matter **8**, 9341 (1996).
- [41] A. Di Cicco, M. Minicucci, and A. Filipponi, *New advances in the study of local structure of molten binary salts*, Phys. Rev. Lett. **78**, 460 (1997).
- [42] S. De Panfilis and A. Filipponi, *Short-range order in liquid tellurium probed by x-ray absorption spectroscopy*, Europhys. Lett. **37**, 397 (1997).
- [43] M. Minicucci and A. Di Cicco, *Short-range structure in solid and liquid CuBr probed by multiple-edge X-ray Absorption Spectroscopy*, Phys. Rev. B **56**, 11456 (1997).
- [44] J. P. Itié, A. Polian, D. Martinez-Garcia, V. Briois, A. Di Cicco, A. Filipponi, and A. San Miguel, *X-ray absorption spectroscopy under extreme conditions*, J. Phys. IV, Colloq. (France) **7**, C2-31 (1997).
- [45] A. Di Cicco, *Short-range structure in solid and liquid matter using multiple-edge EXAFS refinement*, J. Phys. IV Colloq. C2, 171 (1997).
- [46] A. Di Cicco, M. Rosolen, R. Tossici, R. Marassi, and J. Rybicki, *Radial Distribution Function in Solid and Liquid KBr Probed by EXAFS*, J. Phys. IV Colloq. C2, 245 (1997), France.
- [47] M. Giorgetti, M. Berrettoni, R. Marassi, A. Di Cicco and I. Ascone, *XAS Studies on "Costa Type" Model Compounds of Vitamin B<sub>12</sub>*, J. Phys. IV Colloq. C2, 621 (1997), France.
- [48] M. Minicucci and A. Di Cicco, *Multiple-edge EXAFS Study of Solid and Liquid CuBr*, J. Phys. IV Colloq. C2, 1027 (1997), France.
- [49] Y. Katayama, M. Mezouar, J. P. Itié, J. M. Besson, G. Syfousse, P. LeFevre, and A. Di Cicco, *High-pressure high-temperature XAFS investigation of HgTe*, J. Phys. IV Colloq. C2, 1011 (1997), France.
- [50] F. Sperandini and A. Di Cicco, *Local Structure EXAFS Study of Nd<sub>2</sub>CuO<sub>4</sub> and Nd<sub>1.85</sub>Ce<sub>0.15</sub>CuO<sub>4</sub> Using Multiple-Scattering Data-Analysis*, J. Phys. IV Colloq. C2, 1237 (1997), France.
- [51] S. Pascarelli, F. Boscherini, C. Lamberti, and S. Mobilio, *Tetragonal-strain-induced local structural modifications in InAs<sub>x</sub>P<sub>1-x</sub>/InP superlattices: A detailed x-ray absorption investigation* Phys. Rev. B **56**, 1936 (1997).
- [52] A. Filipponi, D. T. Bowron, C. Lobban, and J. L. Finney, *Structural determination of the hydrophobic hydration shell of Kr*, Phys. Rev. Lett. **79**, 1293 (1997).
- [53] M. Giorgetti, M. Berrettoni, A. Filipponi, P. J. Kulesza, and R. Marassi, *Evidence of four-body contributions in the EXAFS spectrum of Na<sub>2</sub>Co[Fe(CN)<sub>6</sub>]*, Chem. Phys. Lett. **275**, 108 (1997).
- [54] U. Buontempo, A. Di Cicco, A. Filipponi, M. Nardone, and P. Postorino, *Determination of the I<sub>2</sub> bond-length distribution in liquid, solid, and solution, by EXAFS spectroscopy*, J. Chem. Phys. **107**, 5720 (1997).
- [55] H. H. Zhang, A. Filipponi, M. J. Scott, R. H. Holm, B. Hedman, and K. O. Hodgson, *Multiple-edge XAS studies of Cyanide-Bridged Iron-Copper Molecular Assemblies Relevant to Cyanide-Inhibited Heme-Copper Oxidases Using Four-Body Multiple-Scattering Analysis*, J. Am. Chem. Soc. **119**, 2470 (1997).
- [56] S. Pascarelli, F. Boscherini, C. Lamberti, and S. Mobilio, *Tetragonal strain induced local structural modifications in InAs<sub>x</sub>P<sub>1-x</sub>/InP superlattices: a detailed x-ray absorption investigation* Phys. Rev. B, **56** 1936 (1997).
- [57] Y. Wang, J. L. DuBois, B. Hedman, K. O. Hodgson, T. D. P. Stack, *Catalytic Galactose Oxidase Models: Biomimetic Cu(II)-Phenoxyl-Radical Reactivity*, Science **279**, 537 (1998).
- [58] K. B. Musgrave, H. C. Angove, B. K. Burgess, B. Hedman, and K. O. Hodgson, *All-ferrous Titanium(III) Citrate Reduced Fe Protein of Nitrogenase: An XAS study of Electronic and Metrical Structure*. J. Am. Chem. Soc. **120**, 5325 (1998).
- [59] F. Boscherini, S. De Panfilis, and J. Weissmüller, *Determination of local structure in nanophase Pd by x-ray absorption spectroscopy*, Phys. Rev. B **57**, 3365 (1998).
- [60] U. Buontempo, A. Filipponi, P. Postorino, and R. Zaccari, *Density effect on molecular bond length in fluid I<sub>2</sub> by x-ray absorption spectroscopy*, J. Chem. Phys., **108**, 4131 (1998).

- [61] U. Buontempo, A. Filipponi, D. Martínez-García, P. Postorino, M. Mezouar, and J. P. Itié, *Anomalous bond length expansion in liquid iodine at high pressure*, Phys. Rev. Lett. **80**, 1912 (1998).
- [62] F. Cardellini, V. Contini, G. D' Agostino, and A. Filipponi, *On X-ray diffraction and X-ray absorption spectroscopy characterization of ball milled iron copper solid solution*, Mater. Sci. Forum (Switzerland) **269&272**, 473 (1998).
- [63] F. Sperandini, A. Di Cicco, and M. Gazda, *Short-range structural properties of Nd<sub>2</sub>CuO<sub>4</sub> and Nd<sub>1.85</sub>Ce<sub>0.15</sub>CuO<sub>4</sub> studied by multiple-scattering EXAFS data-analysis*, Phys. Rev. B **57**, 6067 (1998).
- [64] D. T. Bowron, A. Filipponi, C. Lobban, and J. L. Finney, *Temperature-induced disordering of the hydrophobic hydration shell of Kr and Xe*, Chem. Phys. Lett. **293**, 33 (1998).
- [65] A. Filipponi and P. D' Angelo, *Accurate determination of molecular structures by x-ray absorption spectroscopy*, J. Chem. Phys. **109**, 5356 (1998).
- [66] C. Meneghini and S. Morante, *The active site of the Tetanus Neurotoxin resolved by Multiple Scattering Analysis in X-Ray Absorption Spectroscopy*, Biophysical Journal **75**, 1953 (1998).
- [67] D. T. Bowron, A. Filipponi, M. A. Roberts, and J. L. Finney, *Hydrophobic hydration and the formation of a clathrate hydrate*, Phys. Rev. Lett. **81**, 4164 (1998).
- [68] S. Mobilio and C. Meneghini *Synchrotron Radiation in the study of amorphous materials*, J. Non-Cryst. Solids **232-234**, 25 (1998).
- [69] A. Di Cicco, G. Aquilanti, M. Minicucci, A. Filipponi, and J. Rybicki, *Short-range interaction in liquid rhodium probed by x-ray absorption spectroscopy*. J. Phys.: Condensed Matter **11**, L43 (1999).
- [70] R. E. Benfield, A. Filipponi, N. Morgante, and G. Schmid, *Structural characterisation of the giant organometallic platinum cluster Pt<sub>309</sub>(phen\*)<sub>36</sub>O<sub>30</sub> using EXAFS*, J. Organomet. Chem. **573**, 299 (1999).
- [71] A. Filipponi, M. Borowski, F. Natali, *Multiple scattering x-ray absorption investigation of the local structure around substitutional Ge impurities in Ag*, Phys. Rev. B **60**, 6 (1999).
- [72] A. Filipponi, A. Di Cicco, and S. De Panfilis, *Structure of undercooled liquid Pd probed by x-ray absorption spectroscopy*, Phys. Rev. Lett. **83**, 560 (1999).
- [73] A. Filipponi, A. Di Cicco, G. Aquilanti, M. Minicucci, S. De Panfilis, and J. Rybicki, *Short-range structure of liquid palladium and rhodium at very high temperatures*, J. Non-Cryst. Solids **250-252**, 172 (1999).
- [74] P. Postorino, U. Buontempo, A. Filipponi, and M. Nardone, *Early metallization in molecular fluids: the case of iodine*, Physica B **265**, 72 (1999).
- [75] A. Di Cicco and M. Minicucci, *Solid and liquid short-range structure determined by EXAFS multiple-scattering data-analysis*, J. Synchrotron Rad. **6**, 255 (1999).
- [76] S. Rossano, A. Ramos, J.-M. Delaye, A. Filipponi, S. Creux, Ch. Brouder, and G. Calas, *Iron surrounding in CaO-FeO-2SiO<sub>2</sub> glass: EXAFS and molecular dynamics simulation*, J. Synchrotron Rad. **6**, 247 (1999).
- [77] G. Aquilanti, A. Di Cicco, M. Minicucci, A. Filipponi, and J. Rybicki, *High-temperature EXAFS study of solid and liquid rhodium*, J. Synchrotron Rad. **6**, 251 (1999).
- [78] S. De Panfilis, A. Filipponi, and C. Meneghini, *Local structure in crystalline and liquid Te probed by XAS*, J. Synchrotron Rad. **6**, 549 (1999).
- [79] M. Giorgetti, S. Passerini, M. Berrettoni, and W. H. Smyrl, *XAS investigation on polyvalent cation intercalation in V<sub>2</sub>O<sub>5</sub> aerogels*, J. Synchrotron Rad. **6**, 743 (1999).
- [80] M. Borowski, D. T. Bowron, and S. De Panfilis, *High-energy x-ray absorption at ESRF BM29*, J. Synchrotron Rad. **6**, 179 (1999).
- [81] K. E. Aasmundtveit, E. J. Samuelsen, C. Steinsland, C. Meneghini and A. Filipponi, *EXAFS studies of iodine-doped poly(octylthiophene)*, Synthetic Metals **101**, 363 (1999).
- [82] P. D' Angelo and N. V. Pavel, *Evidence of three-body correlation functions in Rb<sup>+</sup> and Sr<sup>2+</sup> acetonitrile solutions*, J. Chem. Phys. **111**, 5107 (1999).
- [83] M. Tormen, D. De Salvador, M. Natali, A. Drigo, F. Romanato, G. Rossetto, F. Boscherini, S. Mobilio, *Bond length variation in In<sub>0.25</sub>Ga<sub>0.75</sub>As/InP epitaxial layers thicker than the critical thickness*, J. Appl. Phys. **86**, 2533 (1999).
- [84] S. Rossano, A. Ramos, J.-M. Delaye, S. Creux, A. Filipponi, Ch. Brouder, and G. Calas, *EXAFS and Molecular Dynamics combined study of CaO-FeO-2SiO<sub>2</sub> glass. New insight into site significance in silicate glasses*, Europhys. Lett. **49**, 597 (2000).

- [85] R. Laskowski, J. Rybicki, M. Chybicki, and A. Di Cicco, *Pair potential for Solid Rhodium from Extrapolated Ab-Initio Data*, Phys. Stat. Sol. (b) **217**, 737 (2000).
- [86] S. Pascarelli, S. De Panfilis, and T. Neisius, *Local structural rearrangement in deeply undercooled liquid Ge probed by subsecond time-resolved x-ray-absorption spectroscopy*, Phys. Rev. B **62**, 3717 (2000).
- [87] A. Witkowska, J. Rybicki, K. Trzebiatowski, A. Di Cicco, and M. Minicucci, *Influence of hydrogen reduction on the structure of PbSiO<sub>3</sub> glass: an EXAFS study*, J. Non-Cryst. Solids **276**, 19 (2000).
- [88] A. Di Cicco, L. Comez, J. P. Itié, and A. Polian, *High pressure x-ray absorption study of gallium and bismuth*, Proceedings of the AIRAPT-17 conference (1999), to be published.
- [89] A. Di Cicco, M. Taglienti, M. Minicucci, and A. Filipponi, *Short-range structure of solid and liquid AgBr determined by multiple-edge X-ray Absorption Spectroscopy*, Phys. Rev. B (2000), to be published.

#### Applications (subject index)

1. Simple molecules – gas phase  
[5], [8], [9], [12], [27], [54], [65]
2. Complex molecules or molecular solids  
[12], [23], [53], [54], [79], [81]
3. Crystalline solids  
[10], [22], [25], [27], [32], [36], [51] (superlattices), [56] (thin films), [62] (ball mill), [71] (impurities), [77, 78], [83] (thin films), [85, 86], [88, 89]
4. Nanocrystalline systems – Clusters  
[4], [22], [29], [59], [70]
5. HTC superconductors  
[14], [30], [50], [63]
6. Amorphous solids and glasses  
[1], [2], [64], [76], [84], [87]
7. Liquid metals  
[6], [7], [17], [25], [32], [34], [39], [40], [42], [45], [69], [72, 73], [75], [78], [88]
8. Molten salts  
[21], [31], [41], [43], [46], [48], [75], [89]
9. Molecular liquids  
[12], [56], [57], [74]
10. Solutions  
[15], [16], [26] (micelle), [33], [35], [54], [82]
11. Hydrophobic hydration  
[52], [64], [67]
12. High-pressure  
[36], [44], [49], [61], [74], [88]
13. Biological matter  
[13], [18], [19], [20], [24], [28], [37], [38], [47], [55], [57-58], [66]
14. XANES  
[11], [36]

#### Double-electron excitations

##### 1. EVIDENCE IN ABSORPTION

- [1] A. Filipponi, E. Bernieri and S. Mobilio, *Multielectron excitations in X-Ray absorption spectra of a-Si:H*, Phys. Rev. B **38**, 3298 (1988).
- [2] S. Bodeur, P. Millié, E. Lizon à Lugrin, I. Nenner, A. Filipponi, F. Boscherini, and S. Mobilio, *Double core vacancy excited states in the photoabsorption spectrum of SiX<sub>4</sub> (X=H, CH<sub>3</sub>, F, Cl, Br) at the silicon 1s edge*, Phys. Rev. A **39**, 5075 (1989).
- [3] A. Di Cicco, S. Stizza, A. Filipponi, F. Boscherini, and S. Mobilio, *X-ray absorption investigation of SiX<sub>4</sub> (X=Cl, F, CH<sub>3</sub>)*, J. Phys. B: At. Mol. Opt. Phys. **25**, 2309 (1992).
- [4] P. D' Angelo, A. Di Cicco, A. Filipponi, and N. V. Pavel, *Double electron excitation channels at the Br K-edge of HBr and Br<sub>2</sub>*, Phys. Rev. A **47**, 2055 (1993).

- [5] E. Burattini, P. D' Angelo, A. Di Cicco, A. Filipponi, and N. V. Pavel, *Multiple scattering x-ray absorption analysis of simple brominated hydrocarbon molecules*, J. Phys. Chem. **97**, 5486 (1993).
- [6] A. Filipponi, T. A. Tyson, K. O. Hodgson, and S. Mobilio, *The KL-edge in x-ray absorption spectra of third period atoms: Si, P, S, and Cl*, Phys. Rev. A **48**, 1328 (1993).
- [7] A. Filipponi, L. Ottaviano, and T. A. Tyson, *Double-electron excitation channels at the L-edges of atomic Hg*, Phys. Rev. A **48**, 2098 (1993).
- [8] A. Di Cicco and A. Filipponi, *Evidence for [2p(s)4f] multi-electron resonances in x-ray absorption spectra of sixth period elements*, Phys. Rev. B **49**, 12564 (1994).
- [9] R. E. Benfield, A. Filipponi, D. T. Bowron, R. J. Newport, and S. J. Gurman, *X-ray absorption study of gold coordination compounds: EXAFS refinements and double-electron excitation background*, J. Phys.: Condensed Matter **6**, 8429 (1994).
- [10] A. Filipponi, *Double-electron excitation effects above inner shell x-ray absorption edges*, Physica B **208 & 209**, 29 (1995).
- [11] A. Filipponi and A. Di Cicco, *Atomic background in X-ray absorption spectra of fifth period elements: evidence for double-electron excitation edges*, Phys. Rev. A **52**, 1072 (1995).
- [12] A. Filipponi and A. Di Cicco, *Short-range order in crystalline, amorphous, liquid, and supercooled liquid germanium probed by x-ray absorption spectroscopy*, Phys. Rev. B **51**, 12322 (1995).
- [13] A. Di Cicco and F. Sperandini, *Cu-O-Cu in-plane angle distribution in layered superconductors probed by EXAFS*, Physica C **258**, 349 (1996).
- [14] P. D' Angelo, H.-F. Nolting, and N. V. Pavel, *Evidence for multielectron resonances at the Sr K edge*, Phys. Rev. A **53**, 798 (1996).
- [15] A. Filipponi and A. Di Cicco, *Comment on "X-ray absorption fine structure in embedded atoms"*, Phys. Rev. B **53**, 9466 (1996).

## 2. EVIDENCE IN PHOTOEMISSION

- [1] A. Filipponi, S. Di Nardo, M. Passacantando, L. Lozzi, S. Santucci, and P. Picozzi, *1s shake-up x-ray photoelectron spectrum of Na in NaCl and other Na salts*, Phys. Rev. B **48**, 13430 (1993).
- [2] A. Di Cicco, M. De Crescenzi, R. Bernardini, and G. Mancini, *Evidence of KL double-electron excitations in X-ray photoemission spectra of solids: the case of MgO and MgF<sub>2</sub>*, Phys. Rev. B **49**, 2226 (1994).
- [3] A. Filipponi, M. Passacantando, L. Lozzi, S. Santucci, P. Picozzi, and A. Di Cicco, *1s shake-up excitations in NaF, NaCl, NaBr, and Na<sub>2</sub>SO<sub>4</sub>*, Solid State Comm. **91**, 555 (1994).
- [4] A. Filipponi, A. Di Cicco, P. Pianetta, and T. Kendelewicz, *Evidence for [1s2p]3p shakeup channels in compounds and oxides of third period elements*, Phys. Rev. B **53**, 15571 (1996).
- [5] V. Formoso, A. Filipponi, A. Di Cicco, G. Chiarello, R. Felici, and A. Santaniello, *1s shake-up excitations in Ge and GeO<sub>2</sub> by high-energy x-ray photoemission spectroscopy*, Phys. Rev. B **61**, 1871 (2000).

## A3. Package availability, usage and software dependencies

The GNXAS package is presently distributed on the World Wide Web at the L' Aquila and Camerino URLs.

- [1] L' Aquila URL: <http://www.aquila.infn.it/gnxas>, where Linux, SunOS, and HP-UX versions are currently tested and compiled.
- [2] Camerino URL: <http://gnxas.unicam.it>, where Linux, OpenVMS and Digital UNIX are currently tested and compiled.

All updated information regarding codes, documentation, installation will be found on these sites. Full details about characteristics and usage of the main programs of the GNXAS package are given in the specific documentation associated with each program.

The `crymol`, `phagen`, `gnxas`, and `fitheo` programs are designed to run using input command files containing the relevant information. The structure of the input command files is described in details in the documentation of the individual programs.

The typical way to run GNXAS programs under UNIX operating systems is to write a command

```
gnxas < gnxas.in
```

where `gnxas` is a program of the package and `gnxas.in` is the input command file. The output can also be written in a file using the command

```
gnxas < gnxas.in > gnxas.out
```

Under OpenVMS, the same command file `gnxas.in` must also contain the appropriate DCL commands, with at least one extra-line in top of the file:

```
$ run gnxas
```

Usually, the name of the command file contains the extension `*.com` in place of `*.in`. The `gnxas` program can be then run interactively using the DCL command

```
@gnxas.com
```

or using a batch procedure with a given priority (output is sent to a `*.log` file).

Under both operating systems, it is of course necessary to define the relevant path(s) to run the GNXAS programs.

Most of the graphical output of the GNXAS package is written for the TOPdrawer program. This program was originally developed at SLAC and more recently it has been widely supported by a Japanese group (see ref. [21]).

The `fitheo` code makes use of the MINUIT subroutine of the CERN library. Details on the numerous usable commands and on the parts of the `fitheo` input which include MINUIT instructions can be found on the MINUIT long writeup (see ref. [13]).

