GNXAS. II. Structural refinement of experimental data

Adriano Filipponi, Andrea Di Cicco and Calogero Renzo Natoli

* Dipartimento di Scienze Fisiche e Chimiche, Università degli Studi dell’Aquila, Via Vetoio, I-67100 L’Aquila, Italy,
b Physics Division, School of Science and Technology, Università degli Studi di Camerino, Via Madonna delle Carceri, I-62032 Camerino, Italy, and *INFN Laboratori Nazionali di Frascati, Via Enrico Fermi 40, CP 13, I-00044 Frascati, Italy.

*Correspondence e-mail: adriano.filipponi@aquila.infn.it

Fundamental aspects related to the XAFS signal fitting procedure adopted in the FITHEO program of the GNXAS package are reviewed with a particular emphasis on the approach used to treat the configurational averages. Examples of representative applications are illustrated.

1. Introduction

The FITHEO program implements a versatile nonlinear least-squares fitting routine for the modelling of raw X-ray absorption spectra and also includes the possibility of multi-edge fitting. The input is the experimental spectrum (or spectra) and the signals required for modelling as calculated for the reference structure from the PHAGEN/GNXAS suite (Filipponi et al., 2020). Its main features are the following.

(i) The direct application to raw data (absorption versus energy) without the need for preliminary background subtraction or Fourier filtering.

(ii) The possibility of adopting a versatile menu of atomic backgrounds also accounting for double-electron excitation effects.

(iii) The signals are associated with the corresponding peaks of the n-body distributions and are refined with a set of constrained parameters which define coordination, geometry and disorder effects.

(iv) The disorder average is treated by suitable expressions that generalize the EXAFS Debye–Waller factor to multidimensional and nonsymmetrical cases, as described in Section 2.

(v) Statistical confidence intervals for the parameter are obtained in a Bayesian formalism, as described in Section 3.

The approach has been successfully applied to a wide range of structural problems.

2. Disorder models and ensemble average

In the general case, each multiple-scattering (MS) signal [or irreducible pair \( \gamma^{(2)}(r, k) \), triplet \( \gamma^{(3)}(r_1, r_2, \theta, k) \), ... signals] (Filipponi et al., 1995) will be described by an amplitude \( A(k, r) \) and a phase \( \psi(k, r) \) dependent on the wavevector \( k \) and on a set of rotationally invariant coordinates, jointly indicated with \( r \), as

\[ \chi(k) = \sum A(k, r) \exp[i\psi(k, r)]. \]  

(1)
where \( \Im \) indicates the imaginary part. The ensemble average (Filipponi, 2001) over the corresponding pair, triplet, ... distributions is given (for an elemental system) by

\[
\langle \chi(k) \rangle \simeq \int_0^\infty dr \frac{4\pi}{\lambda} \rho(g(r) \chi^{(2)}(r, k)} + \int dr_1 dr_2 d\theta \frac{8\pi}{\lambda^2} \sin(\theta) \rho^2 g_3(r_1, r_2, \theta)
\]

\[
\times \chi^{(3)}(r_1, r_2, \theta, k) + \ldots.
\]

(2)

The integrals can be split into separate integrations over the probability densities \( P_j(r) \) of isolated peaks \( j \), each including the corresponding thermal and static disorder effects consistent with statistical mechanics,

\[
\langle \chi(k) \rangle = \Im \sum_j \int A_j(k, r) \exp[i\psi_j(k, r)] P_j(r) dr.
\]

(3)

For each peak, it is convenient to exploit independent Taylor expansions of \( A(k, r) \) and \( \psi(k, r) \) around a corresponding central reference configuration \( r_0 \) (Benfatto et al., 1989) that can be written, with \( r' = r - r_0 \), as

\[
A(k, r) = A_0(A(k, r))_{r_0} + A_1(A(k, r))_{r_0} \cdot r' + \frac{1}{2}[r', A_2(k, r)]_{r_0} + \ldots
\]

(4)

\[
\psi(k, r) = \psi_0(k, r)_{r_0} + \psi_1(k, r)_{r_0} \cdot r' + \frac{1}{2}[r', \psi_2(k, r)]_{r_0} + \ldots
\]

(5)

to obtain a convenient general expression for the ensemble-averaged signal over a single peak,

\[
\langle \chi(k) \rangle \simeq \Im \int \left[ A_0 + A_1 \cdot r' + \frac{1}{2}(r', A_2) + O(r'^3) \right] P(r) dr.
\]

(6)

It is possible to recast this expression factorizing out the undamped signal and the amplitude expansion as a differential operator

\[
\langle \chi(k) \rangle = \Im A_0 \exp(i\psi_0) \frac{\partial \exp}{\partial \psi_1} \left( \frac{1}{iA_0} A_1 \cdot \frac{\partial}{\partial \psi_1} \Phi_1 \right) + \ldots
\]

(7)

acting on an integral, which resembles the characteristic function of the peak probability density, apart from a modifier term that vanishes when the \( \psi_2 \) term is neglected.

The above general expression can be specialized to particular cases.

2.1. First-order and Gaussian distribution

Retaining only first-order terms, the expression (7) for the averaged signal reduces to

\[
\langle \chi(k) \rangle \simeq \Im A_0 \exp(i\psi_0) \left( \frac{1}{iA_0} A_1 \cdot \frac{\partial}{\partial \psi_1} \Phi_1 \right),
\]

(8)

where \( \Phi_1(\psi_1) = \exp[-\frac{1}{2}(\psi_1, M\psi_1)] \) is the characteristic function of the Gaussian distribution and \( M \) is the related covariance matrix; consequently,

\[
\langle \chi(k) \rangle \simeq \Im A_0 \exp(i\psi_0) \left[ 1 + \frac{i}{A_0} (A_1, M\psi_1) \right]
\]

\[
\times \exp \left[ -\frac{1}{2}(\psi_1, M\psi_1) \right].
\]

(9)

For the case of a single scattering signal the leading phase is \( \psi \simeq 2kR + \psi_0 \) and \( \psi_1 \simeq 2k \). The covariance matrix reduces to

\[
\begin{bmatrix}
A_0 & A_1 \\
A_1 & \mu_2
\end{bmatrix}
\]

Figure 1
Examples of molecular structures (Filipponi & D’Angelo, 1998) where the XAFS signal is dominated by pair contributions. The fitting included the first two terms of the multiple-scattering expansion \( \chi_2 \) and \( \chi_4 \), Br2 (left) and GeCl4 at the Ge K edge (right).
the bond-length variance $M = \sigma^2$ and the dominant exponential term reduces to the usual EXAFS Debye–Waller factor. Equation (9) therefore also provides the direct generalization of the signal damping correction in the Gaussian approximation for irreducible $\gamma^{(0)}(r, k)$ (with $n > 2$) signals or related multiple-scattering signals.

2.2. Pair signals in molecular structures

In the case of pair signals, it is convenient to expand the irreducible $\gamma^{(2)}(r, k)$ signal in the corresponding multiple-scattering expansion involving photoabsorber atom 0 and backscattering atom $i$,

$$\gamma^{(2)}(0, i) \sim \chi_{i0}^{00} + \chi_{i2}^{00} + \chi_{i6}^{0000} + \ldots,$$

(10)

since each contribution has a smoother $r$ dependence of the corresponding amplitude and phases with respect to $\gamma^{(2)}$ and an improved numerical accuracy for the average is obtained. The examples in Fig. 1, taken from Filipponi & D’Angelo (1998), clearly show the high quality of the XAFS signal model spectrum for simple molecular structures such as Br$_2$ and GeCl$_4$ in the gas phase. It can be appreciated that the $\chi_{i0}^{00}$ signal is weak but clearly detectable in these spectra.

2.3. Pair signals with asymmetric distributions

In moderately disordered asymmetric cases a useful and versatile model distribution for the probability of finding one atom at distance $r$ from the photoabsorber is given by the Gamma distribution,

$$P(r) = \frac{2}{\sigma^p \Gamma(p \sigma^2)} \left[ \frac{4(r - R)}{(\sigma \beta)} \right]^{(4/\sigma^2) - 1} \times \exp \left\{ -\frac{4}{\sigma^2} + \frac{2(r - R)}{\sigma \beta} \right\},$$

(11)

where $R$, $\sigma^2$ and $\beta$ represent the three distribution parameters: average, variance and skewness. Equation (11) refers to $\beta - R > -2\sigma$, while $P(r) = 0$ elsewhere. $\Gamma(p)$ is Euler’s Gamma function, calculated for $p = 4/\sigma^2$ and defined on the positive real axis. The resulting averaged signal, obtained using equation (7), is given by

$$\langle \chi(k) \rangle = \Im \left\{ A_0 \exp \left\{ i \psi_0 - \frac{4}{\sigma^2} \left( \ln \left( 1 - \frac{i}{2} \sigma \beta \psi_1 \right) - \frac{i}{2} \sigma \beta \psi_1 \right) \right\} \right\} \times \left[ 1 + \frac{A_1}{A_0} \frac{\psi_1 \sigma^2}{(1 - \frac{i}{2} \sigma \beta \psi_1)} \right] + \left( \frac{A_2}{2A_0} + \frac{i}{2} \psi_2 \right) \sigma^2 \left( \frac{1 - \psi_2 \sigma^2}{(1 - \frac{i}{2} \sigma \beta \psi_1)} \right).$$

(12)

This expression has successfully been applied to high-temperature and highly disordered systems and even to model the first-neighbour distance distribution in liquids. It works better than the cumulant expansion because it is associated with a physically sensible distribution with a single additional asymmetry parameter $\beta$ (the third cumulant is $C_3 = \beta \sigma^2$) with respect to the Gaussian model.

2.4. Signals associated with atomic triplets

The geometrical parameterization of an atomic triplet can conveniently be performed using two bond lengths $R_1$ and $R_2$ and the angle $\theta$ between them (Fig. 2), which are usually associated with physical bonds.

In the Gaussian approximation, the distribution is fully characterized by the average values and a covariance matrix,

$$M = \begin{pmatrix} \sigma^2_1 & \sigma^2_{12} & \sigma^2_{13} \\ \sigma_{12} & \sigma^2_2 & \sigma^2_{23} \\ \sigma_{13} & \sigma_{23} & \sigma^2_3 \end{pmatrix},$$

(13)

with a total of six (or less from symmetric cases) independent parameters $\sigma^2_1$, $\sigma^2_2$, $\sigma^2_3$ and the correlations $\rho_{12}$, $\rho_{13}$ and $\rho_{23}$ defined in general as $\rho_{ij} = \sigma_{ij}/(\sigma^2_i \sigma^2_j)^{1/2}$. An example of the application of the above expression is the case of the $\gamma^{(3)}$ (Br–B–Br) signal in the BB$_3$ molecule (in the gas phase) at the Br K edge, the spectrum of which can be accurately reproduced (Filipponi & D’Angelo, 1998) using two pairs and one triplet signal ($D_{3h}$ symmetric geometry; Fig. 3).

2.5. Triplet signals: the special case of collinear paths

If the equilibrium configuration has a collinear ($\theta = \pi$) geometry (i.e. the angle is distributed around $\theta = \pi$), first-order derivatives of the amplitude and phase with respect to $\theta$ vanish (Fig. 4).

The model distribution has to account for the vanishing solid angle element and the direct generalization of a Gaussian distribution is

![Figure 2](image-url)  
**Figure 2** Geometrical parameterization of a triplet configuration; a possible location for the photoabsorber atom is indicated in red.

![Figure 3](image-url)  
**Figure 3** The planar BB$_3$ molecule and the fit of the corresponding XAFS signal (Filipponi & D’Angelo, 1998).
The fitting approach used in the FITHEO program (Filipponi & Di Cicco, 1995) is to attempt to model the experimental raw absorption coefficient data $\alpha_{\text{exp}}(E)$ with a model signal $\alpha_{\text{model}}(E)$, which includes a pre-edge linear background and a post-edge background and the structural signal.

$$
\alpha_{\text{model}}(E) = \begin{cases} 
aE + b & E < E_1 \\
aE + b + J\alpha_{\text{spline}}(E) + \alpha_{\text{hyd}}(E) & E > E_1 
\end{cases}
$$

The $J$ parameter represents the experimental edge discontinuity and is a suitable scaling factor for the smooth polynomial spline contribution $\alpha_{\text{spline}}(E)$ (with unit amplitude and constrained derivative at the edge $E_{\text{ih}}$) and for the structural model signal $\chi_{\text{model}}(k)$, further reduced by the effective overlap factor $S_0$ and multiplied by a smooth hydrogenic cross section decay $\sigma_{\text{hyd}}(E)$. The $\chi_{\text{model}}$ signal is calculated on a theoretical energy mesh with zero corresponding to the continuum threshold; therefore, an additional energy parameter $E_0$ has to be refined to match the experimental photon energy scale $E$ to the theoretical $E - E_0$. $\chi_{\text{model}}$ obviously depends on all structural parameters describing coordination numbers, reference geometries, covariance matrices and asymmetry parameters, according to the previous schemes. The fitting regions include the pre-edge region $E < E_1 < E_{\text{ih}}$ and the XAFS signal region $E > E_2 > E_{\text{ih}}$ since the threshold region cannot be accurately modelled. Exclusion of energy intervals where experimental or model data are not reliable is also possible.

The experimental data are recorded for a sequence of discrete energy points $E_i$ with a suitable mesh and each data is assumed to be an independent absorption measurement with a random Gaussian error $\sigma_i^2$ around the ‘exact’ value, assumed to be predictable by the model. With these assumptions, the probability of obtaining a given experimental spectrum (Filipponi, 1995) results in an uncorrelated multi-dimensional Gaussian function,

$$
f[\alpha_{\text{exp}}(E_i) \mid p] \propto \prod_i \exp \left\{ -\frac{1}{2} \frac{[\alpha_{\text{exp}}(E_i) - \alpha_{\text{model}}(E_i, p)]^2}{\sigma_i^2} \right\}
$$

In the Bayesian formalism (Krappe & Rossner, 2004) the role of the parameter set $p_i$ (conditions in the above expression) and experimental outcomes $\alpha_{\text{exp}}(E_i)$ can be reverted, and equation (13) can be interpreted as the likelihood function $f[p_i \mid \alpha_{\text{exp}}(E)]$ representing the probability of a certain set of parameter values for the given experimental spectrum, assuming a uniform prior for the parameters. This scheme (Filipponi & Di Cicco, 1995; Filipponi, 1995) leads to the usual nonlinear least-square minimization approach that is useful to find the optimal parameter set, is able to reproduce the experimental data and is able to provide information on the confidence intervals (including uncertainties and correlations) associated with the related non-Gaussian multidimensional distribution.

The FITHEO program is also capable of performing multi-edge XAFS refinements (Di Cicco, 1996) as widely applied in recent years. Multi-edge refinements are particularly useful in multi-component systems where local structural information around different atomic sites can be obtained, better constraining the uncertainties in parameter space. Model-independent structural refinement approaches using reverse Monte Carlo methods within the GNXAS–FITHEO method have been also developed (Di Cicco & Trapananti, 2005). Those methods are especially useful to overcome the limitations of the pair distribution peak decomposition of equation (2), which may be difficult to apply in highly disordered systems.

References