

A computer program for analysing complex bulk XAFS spectra and for performing significance tests

Dogan Paktunc

CANMET, Mining and Mineral Sciences Laboratories,
555 Booth Street, Ottawa, ON, Canada K1A 0G1.
E-mail: dpaktunc@nrcan.gc.ca

LSFitXAFS is a computer program designed to calculate fractions of individual components (*i.e.* individual species, minerals, compounds or chemical environments) making up bulk XAFS spectra. The program performs Gaussian elimination and multiple linear regression techniques to simultaneously solve mass balance equations. In addition, the program contains a number of data reduction and analysis routines including the determination of random and systematic noise in Fourier-transformed spectra and calculation of the goodness-of-fit parameters for different combinations of shells.

Keywords: XAFS; data analysis.

1. Introduction

Existing XAFS data-analysis programs such as *EXAFSPAK* (George & Pickering, 1995), *IFEFFIT* (Newville, 2001), *Artemis* (Ravel, 2002) and *WinXAS* (Ressler, 1997) provide comprehensive data processing and analysis capabilities; however, they are not designed to deal with complex mixed spectra resulting from XAFS contributions from a multitude of sources. Environmental and geological samples are often complex mixtures and require isolation of individual species or mineral components contributing to the bulk XAFS. Foster *et al.* (1998) determined the proportion of two end-member components in the XANES spectra of a mine tailings by a linear least-squares technique. Ressler *et al.* (2000) used a combination of the principal component analysis and least-squares technique to determine and quantify the reference model spectra in an unknown XANES spectra. Wasserman (1997) applied the principal component analysis technique to model complex XANES spectra. In addition, as stated by Lytle *et al.* (1989) and Stern (2001), there are needs in routine XAFS data analysis in terms of defining random and systematic noises in XAFS spectra so that the significance of peaks corresponding to discrete shells in Fourier-transformed spectra can be identified and quantified with confidence. Moreover, analysis of spectra involving multishell fitting require that the significance of shell additions to the fit be determined. The *LSFitXAFS* code was developed to address these needs while analysing complex bulk XAFS data representing mixtures of four or more arsenical minerals in mine tailings (Paktunc *et al.*, 2003, 2004) and synthetic arsenical precipitates (Paktunc & Dutrizac, 2003). There is a plan to incorporate a principal component analysis algorithm in the near future to aid the selection of the end-member components. The program is a 32-bit software written in Visual Basic that runs under Windows 9x/NT/2000/XP platforms.

2. Program components

In essence, the program provides three main functions: (i) quantitative determination of individual components such as individual species, minerals, compounds or chemical environments in a complex XAFS spectra; (ii) goodness-of-fit calculations for assessing the

significance of shell additions to the fit; and (iii) determination of random and systematic noise in Fourier-transformed spectra. These functional components are not sequential and can be run independently as described in the following sections.

3. Least-squares fit to estimate component quantities

3.1. Theory

Fractions of the components making up the bulk XAFS are determined by solving a simultaneous set of mass balance equations,

$$\chi_i^{\text{bulk}}(k) = \sum_{j=1}^m \chi_i^j(k) q^j, \quad (1)$$

where $\chi(k)$ is the bulk XAFS, m is the number of components (*i.e.* minerals or compounds), q is the fraction of component m , and i is the number of experimental data points. Because the system of linear equations is overdetermined (*i.e.* $i \gg m$), Gaussian elimination and multiple regression techniques were employed to reduce the solutions to one. The solution set that minimizes the sum of squares of the difference between the observed and theoretical bulk XAFS as per the above mass balance equation is the best fit.

The fractions should sum to 1 for n number of components as indicated by the equation below, but this is not constrained in the program in order to allow another quantity whereby the goodness of fit can be judged,

$$\sum_{m=1}^n q^m = 1. \quad (2)$$

In this case the best-fit solutions where the totals deviating greatly from 1 are not considered as good fits.

3.2. Operation

A components library file must be loaded first. This is done by clicking the second button on the toolbar or selecting the 'Open Components File' menu item under 'File' (Fig. 1). This is a comma-delimited text file (*i.e.* .csv). The first row contains information about the number of components and number of observation or measurement points in the file. Beginning with the second row, the data are arranged in columns [first column: k ; second and subsequent columns: $\chi(k)$]. Such a file is created by the *LSFitXAFS* utility labelled 'Components' under the 'Process' menu item or by a spreadsheet program (saved as a comma-delimited text file). A sample file (*TestComponentsFile.csv*) composed of four components is included with the *LSFitXAFS* package as an example.

Once the components file is loaded, the disabled 'File Open' button (the third button from left) becomes functional. A bulk XAFS data file can be read by clicking the 'File Open' button or by selecting 'Open Bulk Data File' from the 'File' menu item. This operation transfers the data to the column labeled 'measured' of the grid under 'Bulk EXAFS- k '. This file is also a text file, delimited by commas and composed of two columns of data. The first column is k whereas the second column signifies $\chi(k)$. A sample file (*TestBulkXAFS.csv*) is included as an example.

Components making up the bulk spectra are selected from the drop-down list of components by typing the component name directly in the entry box or by scrolling down the list. As a component name is typed, the program searches the list and highlights it in the box. As soon as a component is highlighted, the 'right arrow' button becomes enabled. When this button is pressed, the component highlighted is selected and placed in the box on the right. Components can be removed from the selected components list by highlighting and pressing the 'left arrow' button. New components can be added to the

list through the components utility under the 'Process' menu item. Because the number of XAFS 'point' measurements are high, there is virtually no limit to the number of components selected which would compromise the determination.

When the 'Calculate' button is pressed, results are displayed under the 'Bulk EXAFS-k' and 'Least-squares fraction' grids (Fig. 1). Note that the 'Calculate' button is disabled when there is no component selected. The 'fit' column is the estimated bulk XAFS based on the calculated weight fractions for each component. The 'residual' column is the difference between the observed (measured) and estimated (fit) bulk XAFS. The results listed under 'conc.' are fractions as calculated and under 'fraction' are fractions normalized to 1. As indicated earlier, the calculations do not constrain the total to 1 as an additional measure of the goodness of fit or indicator of error made in the selection of components. Results with sum of weight fractions deviating significantly from 1 should be considered with care. This is listed below the grid as the 'Deviation from optimal total'. In addition, fractions are not constrained to positive numbers either. Negative numbers indicate that either the component with the negative mode is not present or there is incompatibility among the components selected.

The 'Residual sum of squares' value below the box is the sum of the squared values of the 'residual' column. The smaller the difference between the fit and measured values, the better is the fit. 'Sample variance', which is the squared standard deviation, is another measure of the goodness of the fit. If the results are not acceptable, the user can revise the list of components selected by adding or removing and recalculating the modes with the new list.

The results along with the measured values can be viewed by clicking the 'Plot' button.

3.3. Components file

A new components file can be created or additional components can be added to an existing one by the utility 'Components' under the 'Process' menu item. When the 'Read New Component' button is pressed, a two-column comma-delimited text file [first column: k ; second column: $\chi(k)$] is loaded. The file name appears in the input box above. The name should be edited for easy recognition of the component in the file. When the 'Normalize' button is pressed, the values are normalized according to the minimum, maximum and step-size parameters defined in the adjacent boxes.

4. Goodness-of-fit calculations

4.1. Theory

As discussed by Thompson *et al.* (1997), analysis of spectra involving multishell fitting requires that the significance of shell additions to the fit must be determined. Beginning with the second

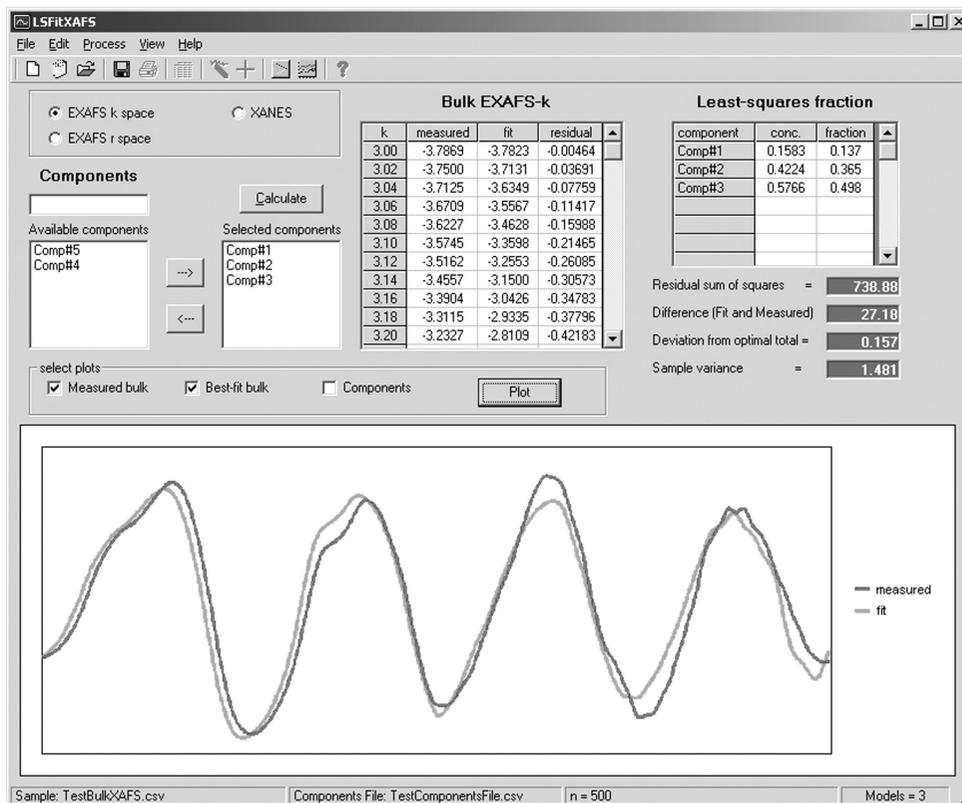


Figure 1 Main window of LSFITXAFS with an example of least-squares fitting.

shell, a relative goodness-of-fit parameter is calculated for each shell addition to test the presence or lack of improvement to the fit. The goodness of fit is the reduced χ^2 value (ϵ^2),

$$\epsilon^2 = \frac{N_{\text{ind}}}{(N_{\text{ind}} - p)n} \sum_{i=1}^n w_i [\chi^{\text{exp}}(k_i) - \chi^{\text{fit}}(k_i)]^2, \quad (3)$$

where N_{ind} is the number of independent data points, p is the number of parameters floated in fits, n is the number of experimental data points, w_i is the weighting parameter, and $\chi^{\text{exp}}(k)$ and $\chi^{\text{fit}}(k)$ are the experimental and fitted XAFS functions. N_{ind} can be defined as

$$N_{\text{ind}} = a(\Delta k \Delta R / \pi) + b, \quad (4)$$

where Δk is the range in k space and ΔR is the range in R space used in fitting, and a is either 0.5 or 2 and b is an integer whose value ranges from -1 to $+2$ as suggested or used by various researchers (*e.g.* Lytle *et al.*, 1989; Stern, 1993, 2001; Stern *et al.*, 1995; Thompson *et al.*, 1997; Bunker, 2002; Finch *et al.*, 2003). In agreement with the commonly used formula for N_{ind} , the default value of a is 2 in the program. According to Stern (1993, 2001), the value of b should be 2. This provides two additional degrees of freedom which means that more parameters can be floated during fitting. The default value of b is 0 in the program, which is conservative for the number of independent data points as it restricts the number of floating parameters. Because background subtraction removes information at low R , ΔR is defined as the range over which there is real data (*i.e.* 0.8 \AA as the starting R value).

The weighting parameter (w_i) can be defined in three ways. The first one is the no-weighting option where $w_i = 1$ (see Stern *et al.*, 1995). The second option assumes a statistical weighting where $w_i = 1/[\chi^{\text{exp}}(k_i)]^2$. In the third option, the weighting (w_i) is defined as

the average error bar representing three times the standard deviations from individual measurement points. This is referred to as the fixed-error-bar weighting. *LSFitXAFS* also supports a fourth parameter whereby the goodness of fit can be assessed. This is the $\Delta\chi^2$ parameter as defined by Stern *et al.* (1995) and 'figure of merit' as per Manceau *et al.* (2000),

$$\Delta\chi^2 = \frac{\sum_{i=1}^n [\chi^{\text{exp}}(k_i) - \chi^{\text{fit}}(k_i)]^2}{\sum_{i=1}^n [\chi^{\text{exp}}(k_i)]^2}. \quad (5)$$

4.2. Operation

When the 'Read' button is pressed, a text file composed of three columns of data [first column: k ; second column: measured $\chi(k)$; third column: fitted $\chi(k)$] is loaded. The parameters in the input boxes on the right should be entered or edited. If the data is already weighted, the k -weighting parameter should be defined as 1. The number of parameters floated during the fit should be specified in the entry box. Upon pressing the 'Calculate' button, the results are listed in the dark boxes below (Fig. 2). The results should be recorded, new sets of XAFS fits resulting from the addition of different shells should be read and subsequent calculations are performed. The number of experimental data points is read from the fit file. If this number is greater than the original number of experimental data points, it should be corrected in the 'number of experimental data points' entry box.

5. Random and systematic noise in Fourier-transformed spectra

5.1. Theory

Following Stern (2001), the random noise can be defined as the background ripples in the Fourier-transformed spectra. Similar to the detection-limit concept, the random noise can be determined by

measuring the background numerous times and calculating a confidence limit for the background measurements. In *LSFitXAFS*, the random noise is defined as the mean Fourier transform magnitude plus three times the standard deviation in the R range chosen. For individual scans of four, this corresponds to a confidence level of 98.5%. The random noise is an important parameter and should be defined for average and individual scans making up the average to verify the significance of peaks in the Fourier-transformed spectra that are supposed to correspond to frequencies for individual atomic shells (Fig. 3). This is needed to assess whether or not a peak in the Fourier-transformed spectra is representing a shell or simply a background ripple. The background level in this program is defined as

$$\text{FT}[\chi(k)]_{\text{bg}} = \text{FT}[\chi(k)]_{\text{av}} + 3\text{FT}[\chi(k)]_{\text{sd}}, \quad (6)$$

where $\text{FT}[\chi(k)]$ is the Fourier-transformed XAFS, bg is the background, av is the average and sd is the standard deviation. The average and standard deviation are calculated for individual scans within a specified R range, usually at high R values where there is no apparent peak representing an atomic shell (Fig. 4). The variance is defined as

$$v = \left(\sum_{i=1}^n \{ \text{FT}[\chi(k)]_i - \text{FT}[\chi(k)]_{\text{av}} \}^2 \right) / d, \quad (7)$$

where i is the scan number, n is the total number of scans and d is the degrees of freedom ($n - 1$).

Systematic noise could result during data acquisition from effects such as imperfections in the monochromator crystal, detector alignment, scattering, and sample heterogeneity, thickness, concentration and degradation during analysis (Lytle *et al.*, 1989). In addition, systematic errors could arise from data analysis such as the changes in experimental conditions, background removal and normalization. Identification of systematic noise such as apparent glitches in the EXAFS spectra is rather straightforward. Weak systematic noise that

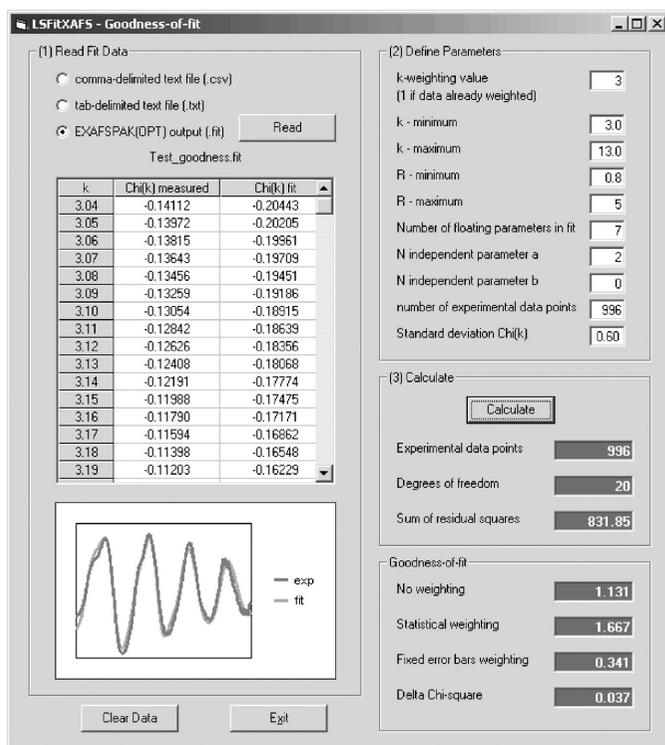


Figure 2
An example of the goodness-of-fit analysis.

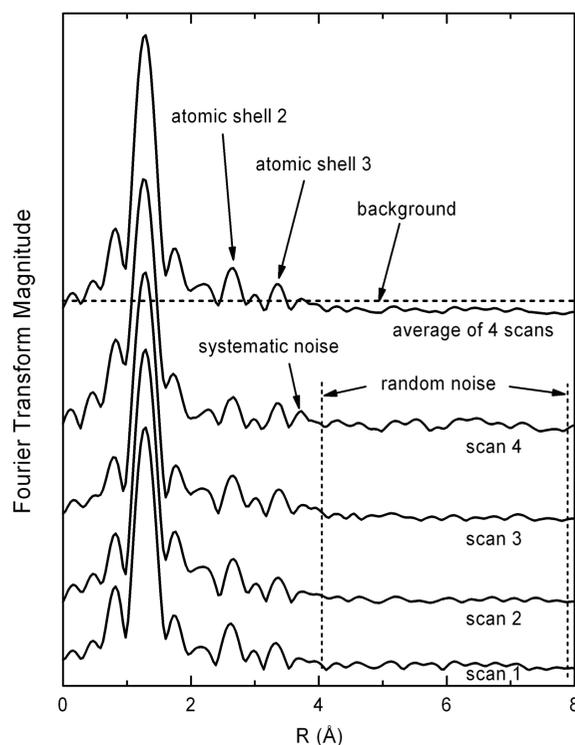


Figure 3
Random and systematic noise in Fourier-transformed spectra.

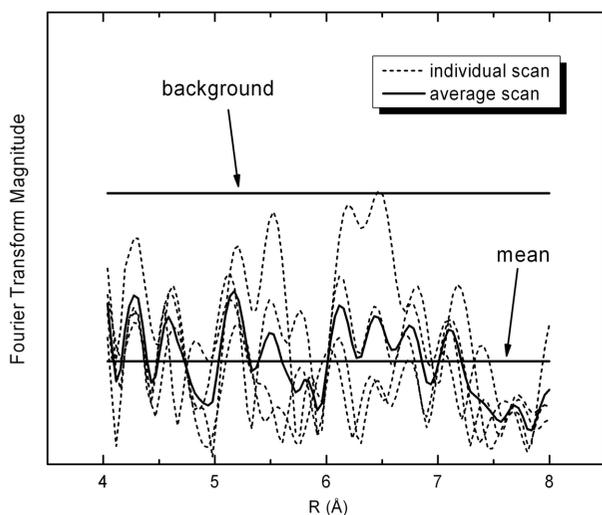


Figure 4
Random noise in Fourier-transformed spectra as defined within the range from 4 to 8 Å. Background is defined as mean plus three times the standard deviation.

remains unrecognized in k space can show up in R space and can be detected in individual scans making up the average scan (Stern, 2001) (Fig. 3). Weak spectral noise that is present in all of the scans would be difficult to recognize.

5.2. Operation

This utility is accessed through the 'Process' menu item. Following selection of the file type, individual scans are read by clicking the 'Read File' button. The scan files are composed of two columns where the first column is R in Å and the second is Fourier-transform magnitude. Because individual scans are normally collected and processed in sequence under identical conditions, the R values are assumed to be identical in each scan. Define the lower and upper R limits in Å in the input boxes. This range should be selected so that there is no obvious major peak representing a shell. Once the 'Calculate' button is pressed, background levels for each scan are calculated and displayed in the grid (Fig. 5).

The Fourier-transform peaks that are above the background level (i.e. random noise) are identified and listed in the grid for each scan. These peaks are considered significant and most likely correspond to atomic shells. When the 'peaks avg' button is pressed, the R values of each peak are averaged and listed in the grid along with their corresponding variance values. If needed, individual scans can be deselected and excluded from the calculation of the average R values.

6. Program requirements

LSFitXAFS is a freeware program written in Visual Basic. The program requires a 32-bit PC with Windows 95, 98, Windows NT, 2000 or XP windows operating system to run. The executable file 'LSFitXAFS.exe' is a stand-alone file and requires VB 6.0 Runfiles to operate. If not present in the system, these files are automatically installed by the set-up package provided.

7. Conclusions

LSFitXAFS is a freeware program designed to quantify the contributions of individual phases to a bulk XAFS spectra and provide useful procedures that are often needed in data analysis but not yet

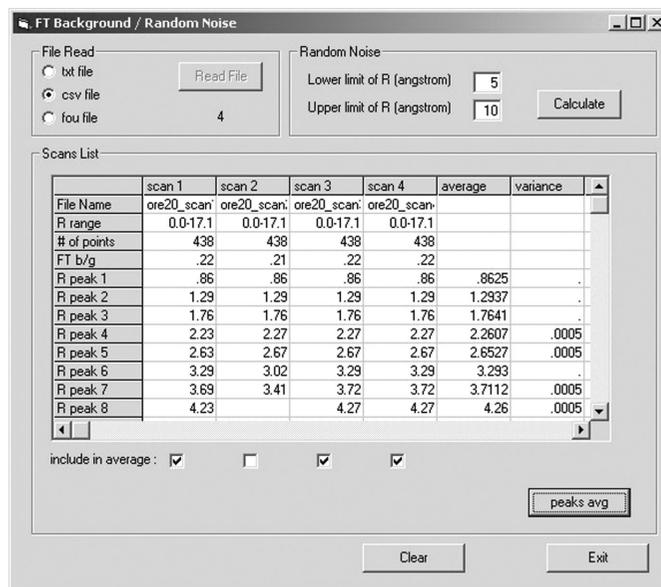


Figure 5
An example of the Fourier-transform random-noise determination.

available as part of the existing XAFS software. The program has been tested and successfully used in the analysis of complex XAFS data from mine tailings and synthetic laboratory precipitates.

The author thanks Andrea Foster for useful discussions on the subject, Allen Pratt, Bill Howell and two anonymous journal referees for the review of the manuscript, and Jeff Cutler for the evaluation of the software.

References

Bunker, G. (2002). *XAFS tutorial documents: data analysis overview*, <http://gbxafs.iit.edu/training/tutorials.html>.

Finch, A. A., Allison, N., Sutton, S. R. & Newville, M. (2003). *Geochim. Cosmochim. Acta*, **67**, 1189–1194.

Foster, A. L., Brown, G. E. Jr, Tingle, T. N. & Parks, G. A. (1998). *Am. Miner.* **83**, 553–568.

George, G. N. & Pickering, I. J. (1995). *EXAFSPAK: a Suite of Computer Programs for Analysis of X-ray Absorption Spectra*. Stanford Synchrotron Radiation Laboratory, Stanford, CA, USA.

Lytle, F. W., Sayers, D. E. & Stern, E. A. (1989). *Physica*, **158**, 701–722.

Manceau, A., Schlegel, M. L., Musso, M., Sole, V. A., Gauthier, C., Petit, P. E. & Trolard, F. (2000). *Geochim. Cosmochim. Acta*, **64**, 3643–3661.

Newville, M. J. (2001). *J. Synchrotron Rad.* **8**, 322–324.

Paktunc, D. & Dutrizac, J. (2003). *Can. Miner.* **41**, 905–919.

Paktunc, D., Foster, A., Heald, S. & Laflamme, G. (2004). *Geochim. Cosmochim. Acta*, **68**, 969–983.

Paktunc, D., Foster, A. & Laflamme, G. (2003). *Environ. Sci. Technol.* **37**, 2067–2074.

Ravel, B. (2002). *ATHENA: an Interactive Graphical Utility for Processing EXAFS Data*. Unpublished.

Ressler, T. (1997). *WinXAS*, <http://www.winxas.de>.

Ressler, T., Wong, J., Roos, J. & Smith, I. L. (2000). *Environ. Sci. Technol.* **34**, 950–958.

Stern, E. A. (1993). *Phys. Rev. B*, **48**, 9825–9827.

Stern, E. A. (2001). *Avoiding Some Pitfalls in XAFS Analysis. Advanced Methods and Tricks of EXAFS Data Modeling Workshop*. Annual NSLS Users' Meeting, 23 May 2001. Abstract.

Stern, E. A., Newville, M., Ravel, B., Yacoby, Y. & Haskel, D. (1995). *Physica B*, **208/209**, 117–120.

Thompson, H. A., Brown, G. E. & Parks, G. A. (1997). *Am. Miner.* **82**, 483–496.

Wasserman, S. R. (1997). *J. Phys. IV*, **7(C2)**, 203–205.