Synchrotron Data Analysis Using XOP

Manuel Sánchez del Río

European Synchrotron Radiation Facility. BP 220 F-38043 Grenoble Cedex (France).

Abstract

A synchrotron experiment goes through several steps: *preparation* (have an idea, think a project, discuss with colleagues, write the proposal, prepare the samples, travel to a facility, perform the safety training…), *realization* (mounting and preparing sample, perform measurements, view results, optimize the measurement, have sleepless nights, etc.) and *home-work* (full analysis of data, extracting results, discuss with colleagues, preparing conclusions, filling user reports, writing papers, travelling to conferences, etc.). A fundamental part in this chain is data analysis, performed i) *on-line* (at the beamline) for improving the measurements and extracting as much of results as possible), and ii) *off-line* (careful detailed analysis at home). This chapter is divided in three parts. In a first part some general ideas of what data analysis means for synchrotron researchers are discussed. In a second part, there is an introduction to basic principles and data analysis for two high impact techniques: diffraction using 2D detectors and X-ray Absorption spectroscopies. A third part is of practical interest, aiming to perform the data analysis operations on real data. For that, it is introduced XOP, a freely available software toolbox for synchrotron applications.

Data analysis in synchrotron research

Main concepts on Data Analysis.

The terminology associated to data analysis has grown in complexity in the last years as a result of the impact of computer science in the simulations and numeric treatment of scientific data. Examples of this avalanche of terminology are these keywords: *Computational science, data analysis, data exploration, data modeling, data analysis, scientific computing, scientific simulations, modeling and simulations,*

computer graphics, numerical simulations, optimization, scientific programming, computational science (physics, chemistry, etc.), imaging, 3D rendering, hyperspectral data, supercomputing, distributed computing, evolutionary computing, etc. In this paragraph I give some general definitions of the different subsets or procedural steps in data analysis.

Data acquisition is the sampling of the real world to generate data that can be manipulated by a computer. Data acquisition typically involves acquisition of signals and processing the signals to obtain desired information. The components of data acquisition systems include appropriate sensors that convert any measurement parameter to an electrical signal, then conditioning the electrical signal which can then be acquired by data acquisition hardware.

Data recording, or the storage of recorded data to backup (disk files) is usually included in data acquisition.

Data analysis is a process of gathering, modeling, and transforming data with the goal of highlighting useful information, suggesting conclusions, and supporting decision making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains. Some actions integrated in the data analysis are: *data reduction* (reduce the amount of data to useful data), *numeric simulation* (describe a phenomenon with a numeric model, usually an equation with some adjustable parameters), *model fitting* (plug a numeric simulation to data), use of *numeric methods* (apply mathematical transformations to numerical data), *ab-initio modeling* (describe a phenomenon from first principles, usually involving high computational work).

Data visualization is the study of the visual representation of data, defined as information which has been abstracted in some schematic form, including attributes or variables for the units of information. Data visualization related to different fields such as information graphics, information visualization, scientific visualization and statistical graphics.

Scientific visualization is an interdisciplinary branch of science, primarily concerned with the visualization of data from multi-dimensional phenomena, such as architectural, meteorological, medical, biological, chemical, physical or geological

systems. Scientific visualization focuses on the use of computer graphics to produce images which aid in understanding of complex scientific concepts. It includes other fields like computer animation, information visualization, surface and volume rendering. From another point of view, the aim of scientific visualization is the creation of pictures from data that help the humans to understand the data behavior.

Data resulting from a Synchrotron Radiation experiment

The scientific activity covered by synchrotron radiation is multidisciplinary, and involves research in different fields like Physics, Chemistry, Medicine, Biology, Geology, etc. and a combination of them. Therefore, many of the concepts concerning data analysis in each field also apply to synchrotron radiation research.

The *beamline* can be conceptually reduced to a data producer, consisting in several scanning parameters (like moving motors of heating or cooling to scan temperature) and several recording parts (signal monitors, detectors, cameras, etc.). Additional parameters that are constant or "slowly" changing, like date, time, ambient temperature, storage ring intensity and emittance, notes or comments from the experimentalists, etc. that could add some valuable information are sometimes called meta-data. During the amount of time that a beamline is dedicated to obtain data from a particular research project (i.e., an *experiment***)**, the beamlines make *measurements* and produce *(numerical) data*. These data are stored *(data storage)* in a computer, usually in the form of independent disk files, but sometimes in a single file that encapsulate a large quantity of data, or in databases. These files are conserved in a server, copied ("backup"), and distributed to other storage devices or computers. They may travel by the network, in a suitcase (external hard drivers), or in the pocket, handbag or even in a necklace (USB keys). The organization of the data into files (or into a single structured file) for further manipulation follows the *data model.* The particular way of arranging data within a file, and how this information of stored in the disk file is called *data format* or *file format.*

Data items resulting from a synchrotron radiation experiment can be:

- Independent variables (abscissas, scanning variables, etc.), usually parameters that gets the information of something that change or move.

- Dependent variables: a signal monitor, intensity counter, detector, etc. which indicates the value of a measurement.
- Additional information (metadata).

For typical 1D (one-dimensional) data, like a single spectrum or diffractogram, one parameter is scanned (independent variable, like beam energy, incident angle, etc.) and a monitor measures a signal that depends on it (depending variable). Sometimes the independent variable is a sequence of indices, i=0,1,2,…. Obviously, this index sequence could be eliminated, giving only the dependent variable array (intensities). This is the case of a "channel analyzer" that gives a number of measured points, one per channel. Any beamline is equipped with many parts that "move" (like motors), or can be scanned (like temperature, time, etc.). Usually many detectors are installed. Therefore, 1D data can be "multi-column" including several independent (scanning) and dependent (monitor) parameters arranged in a multicolumn matrix, each column corresponding to a given parameter, and each row contains all the parameters of a single measurement. This is sometime called n-tuple. N-tuples are easily manipulated using spreadsheet programs, like the spreadsheet in OpenOffice, or its popular ancestor Microsoft Excel.

Typical 2D data are scientific images, where a dependent variable, or monitored signal (pixel value, intensity, I_{ij}) depends on two independent variables (equally gridded and sorted), the pixel coordinates *(i,j)*. One can also build 2D data with a pile or stack or 1D data, arranged either in rows or columns. One can also define some physical parameters associated to the pixel counters *i* and *j*, *x(i)* and *y(i)*. Additionally, a scan that depends on two independent variables (sometimes called *mesh* scan) can be represented as 2D data, if, and only if, the independent variables map a regular grid. 2D data can be represented as a *surface*, as *contour curves* or as an *image*. Images can be monochrome, with grey levels (or darkness) proportional to the intensity value, or use *false colour*, that means that the intensity values are mapped to a colour table, resulting into a coloured image in which colours differ from those that a faithful full-color photograph would show. The colour mapping can cover the full (dynamic) range of intensities, or only a part, and can be linear or non-linear. The relation intensity-value

colour-value gives a large possibility of different visual results for the same data (Figure 1). Tables of intensity values that depend on two coordinates which are not over an homogeneous grid (i.e., they are over scattered *(x,y)* pairs) can only be represented as 2D images after a process of converting the irregular grid onto a regular one (by triangulation, usually Delaunay triangulation) and data interpolation to this new grid.

Data model and file formats from an experiment at ESRF

During a synchrotron radiation experiment an amount of numerical data is produced. The volume of the data depends on the nature of experiment, detectors used, etc. Some experiments produce a small quantity of data that may fit into an old diskette, or a USB key, for example conventional powder diffraction, or magnetic scattering.The use of 2D detectors increases the volume of data, and backup of raw is usually done in DVD's or external hard drives. Data are generally arranged in several files. A *data model* is an abstract model that describes how data is stored. This information is essential for accessing the data in the data analysis chain. Data models formally define data elements or data files, formats, and relationships among them. The use of a well defined data model is essential for an optimizing processing and data analysis, specially when recursive operations should be done on different data sets. Automation of such operations is highly simplified using a data model, with the consequent benefit in time and amount of work for the experimentalist. The simplest and less sophisticate data model is the experiment log-book, where experimentalists write manually the association of data file names to the information related to experimental conditions, sample, sample environment, etc. The use of a log-book as data model implies a hard work for the experimentalist to access, load, visualize and analyze data. These operations can be efficiently automated by software that access an data model structured and written in an adequate form.

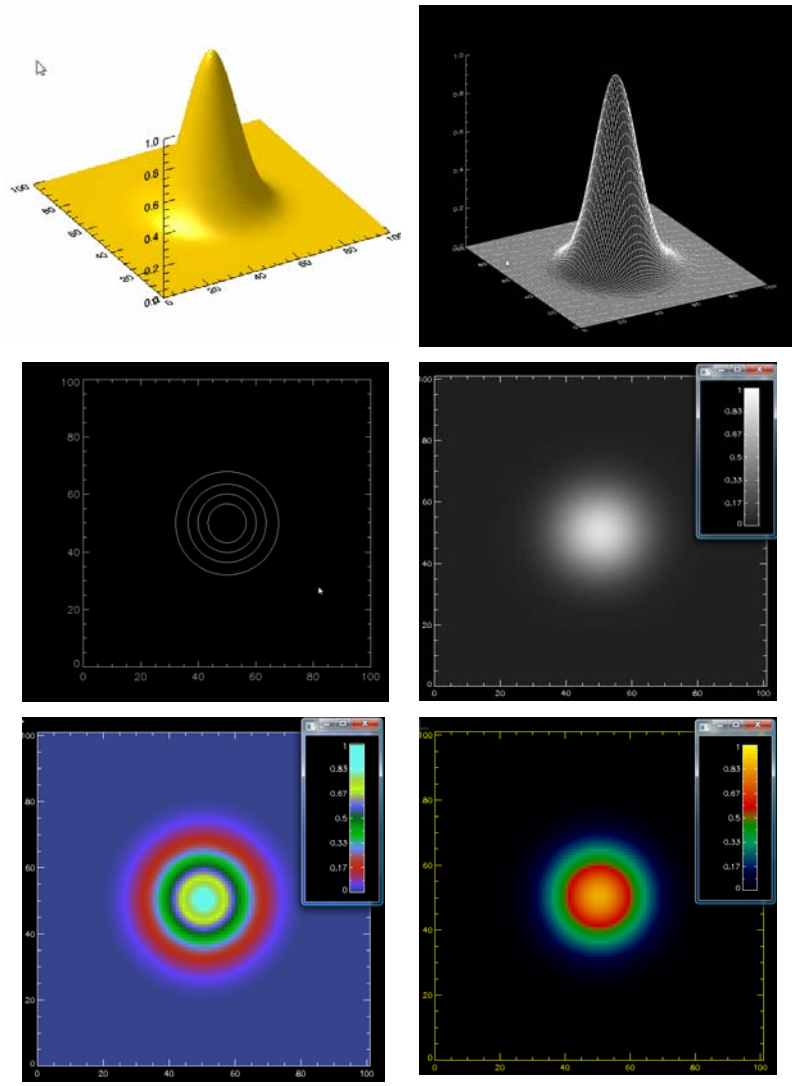

Fig 1: Different representations of a 2D Gaussian. They are generated with the XOP macro shown in Appendix1. Note that the visual look changes drastically depending on the selected color table. Therefore, it is essential to display the colour bar when using false colour representation.

At ESRF different beamlines present different degree of complexity and structuration of the data. Many beamlines use the rudimentary approach (the log-book as data model), and data is stored in files with some arbitrary names. Here, the

relationship between the data and the information associated with these data is *only* in the log-book. On the contrary, other beamlines present a high level of sophistication in the data recording, storage and automated analysis. An example of beamline automation and on-line data analysis for protein crystallography beamlines is the EDNA project (http://www.edna-site.org).

Although each beamline is an independent ecosystem (or compu-system), the data at the ESRF are stored in "more or less" standard files format. As a rule of thumb, data resulting for measurements that produced a limited amount of information are stored in ASCII files. Data from measurements that produce a high amount of data (images, etc.) are stored in binary files. Large data files may be compressed, to save disk space.

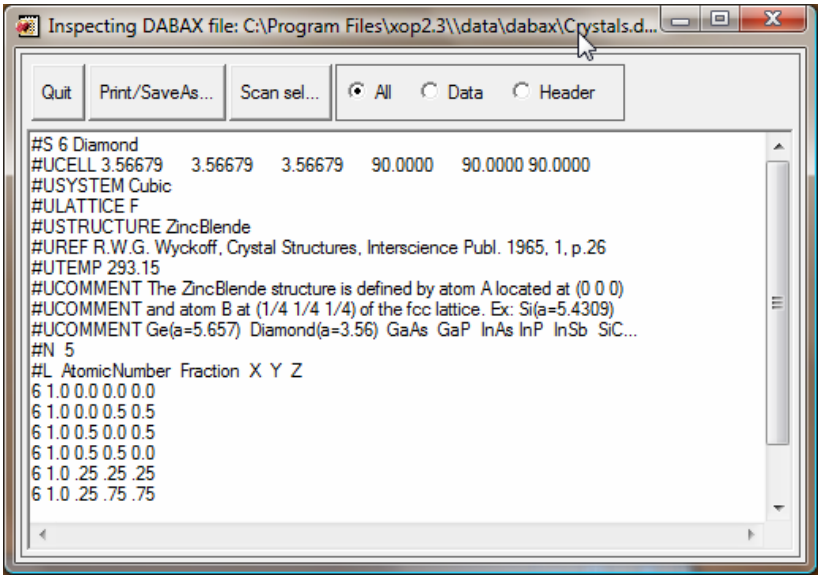

Figure 2 An example of a data set ("scan") in a spec formatted file. The header must contain, at least, three keyword lines starting with #S, #N and #L, for the scan number and title, number of columns and column labels, respectively. The keywords starting with "#U" are optional (U for User defined).

A family of ASCII files used at many ESRF beamlines are somewhat structured and respond to what is called *SPEC file*, because it is the standard way in which the

data acquisition programs SPEC (http://www.certif.com/), using at most of the ESRF beamlines, writes data. The SPEC file is structured in different sequential blocks, called "scans". A single file can contain many scans. Each scan is formed by a header and a data block. The header contains some keywords to define the scan title and number, characteristics of the data block and possibly a number of other parameters (data or metadata) associated to the scan data. The data block is a multicolumn matrix, where the independent and dependent variables are arranged as columns and the number of scanned points (measurements) in rows. The SPEC file may also contain a file header, with information applicable to all scans stored in the file. An example of a minimalist SPEC file is in Figure 2. A different scan format is called MCA (for the initials of Multi-Channel Analyser) where instead of storing data in a multicolumn matrix-like format, the data are stored in many lines of a fixed list of integer numbers (channels) followed by backslash ("\"). SPEC data format is a simple ASCII format able to handle several multicolumn data sets (scans) plus some additional information (keywords in the headers) in a single file.

Binary files are often written at the ESRF when the amount of data is large and the use of ASCII files is inefficient. Although there are many types of binary files that could be used, the choice at the ESRF has been to "invent" a new binary file format, EDF (ESRF Data File) with the following basic requirements (among others): i) ability to store images with different kind of pixel data, like different integers, float, double precision, etc. , ii) It contains an ASCII header that can be read using any file editor and may contains information data from the experiment (metadata), and iii) the data block is binary, and can be easily extracted adapting existing programs or with a few lines of code in several programming languages.

3D data, such as for tomography, are commonly stored in stacks of 2D files, usually stacks of EDF files.

Compression is not used within an EDF file. The files may be compressed using well-known tools (such as gzip). It should be noted that always that compression is applied, it must use lossless compression algorithms that are reversible so that the reconstructed (uncompress) data is identical to the original data. On the contrary, lossy

schemes accept some loss of data in order to achieve higher compression (like in JPEG images) and are not acceptable for scientific data compression.

The idea of putting data into a single (or reduced number) of structured data files, perhaps using a coded data model, has driven to use sophisticated scientific data formats, which are popular in some research fields. For instance FITS (Flexible Image Transport System) (http://fits.gsfc.nasa.gov/)is well used in the astrophysics community. HDF (Hierarchical Data Format) (http://www.hdfgroup.org/)is an open-source project for i) a versatile data model that can represent very complex data objects and a wide variety of metadata, ii) a portable file format with no limit on the number or size of data objects in the collection, and iii) a software library that runs on a range of computational platforms implemented as high-level Application Programming Interface (API) callable by several programming language interfaces. Nexus (http://www.nexusformat.org) is a HDF file format using a particular data model and recommended dictionary of terms (keywords) proposed for the neutronic community and being accepted at present by several synchrotron facilities.

Software programs in the DA chain: from experimental data to publication

The main purpose of a synchrotron experiment is to end with some new results that could contribute to the advancement of science via a published paper. From the origin (experiment) to the end (publication) several steps in the data analysis chain involve the use of computer programs. Depending on the level of abstraction, required preprocessing of the data, additional modeling, volume of the data, etc. several computer application can be used. Sometimes there are general, and perhaps provided by the experimental facility, sometimes are commercial end applications, and other times are customized as-hoc programs written, possible, in several programming languages. Some classifications can be made depending on different criteria:

Looking at the "timing" in the chain:

- *Acquisition software:* ensure the recording of data from experimental apparatus
- *On-line visualization and analysis:* ensure the quality of the measurement, the quality of the data and often a preliminary data analysis,

consisting in basic operations to extract the wanted parameters (background removal, signal normalization, signal transformation, integration, etc.). On-line tools give feedback on the experiment in progress of the experiment, help on deciding the continuation of the measurement and provide information on the subsequent stages and help to optimize the tuning parameters.

- *Off-line data analysis.* Full data analysis on the recorded data usually done after the data has been taken. It may first repeat the preliminary data analysis performed in the On-line phase, perhaps with improved input parameters. It will include full data analysis, data modeling, etc. processes for obtaining the new results. The software tools for data analysis will permit a publication-quality graphics, or at least a good quality graphics and the possibility to export data to other graphing or desktop publishing software.

The software packages or applications can be classified as:

- End applications with a wokspace or visual (or graphical) user interface (VUI, or GUI). Typical examples are Excel, Origin, Igor, Statistica, GNUplot, etc.
- Application programs (high level programming). These are tools that allow the user to write his/her own program for a particularized data treatment or analysis. They contain high level tools, libraries and functions that permit to implement complex algorithms with a very reduced set of computer instructions. Commercial examples: Matlab (http://www.mathworks.com/), Mathematica (http://www.wolfram.com), IDL (http://www.ittvis.com/). Examples of free applications:

  Python+NumPy (http://www.python.org/),

  Scilab (http://www.scilab.org/),

  Octave (http:// www.gnu.org/software/octave/), etc.
- Application tools and libraries. Collection of software code that, once integrated in a user program, allows to create data analysis applications. Typical examples are the numerical libraries, such as:

Numerical Recipes (http://www.nr.com/),
IMSL (http://www.vni.com/products/imsl/),
NAG (http://www.nag.co.uk/), etc.),
Libraries for visual interfaces (Qt (http://www.qtsoftware.com/),
Tk (http://www.tcl.tk/), etc.),
graphical libraries (OpenGL (http://www.opengl.org/),
Pgplot (http:// www.astro.caltech.edu/~tip)

- Application programs (Low level) created in compiled low level programming languages, like C or Fortran. They are used for the more computationally-intensive aspects of scientific computing.

Looking at underlying mathematical aspects the software tools may be divided in:

- Numerical (Matlab, IDL)
- Algebraic (Mathematica, etc.)
- Statistical (Statistica (http://www.statsoft.com/), R (http://www.r-project.org/), etc.)
- Visualization (GNUplot (http://www.gnuplot.info/), Origin (http://www.originlab.com/), Kaleidagraph (http://www.synergy.com/), VTK (http://www.vtk.org/))

Looking at the type of experimental application:

- Generic, multi-disciplinary (visualization, simple maths) (Excel, GNUplot (http://www.gnuplot.info/), ImageJ (http://rsbweb.nih.gov/ij/)).
- Generic for synchrotron radiation at ESRF (PyMCA (http://pymca.sourceforge.net/),XOP (http://www.esrf.eu/UsersAndScience/Experiments/TBS/SciSoft/xop2.3/)
- Targeted applications for some particular fields:

  Examples: powder crystallography

  GSAS (http://www.ccp14.ac.uk/solution/gsas/),

Fullprof (http://www.ill.eu/sites/fullprof/),

FOX (http://vincefn.net/Fox/FoxWiki), etc.)

Examples: X-ray absorption spectroscopy.

XOP/XAID,(http://cars9.uchicago.edu/~ravel/software/),

GNXAS(http://gnxas.unicam.it/XASLABwww/pag_gnxas.html),

MSXAS, FEFF (http://leonardo.phys.washington.edu/feff/),

Examples:X-ray fluorescence spectroscopy.

PyMCA

Basic concepts in Data Analysis for Synchrotron Powder Diffraction and x-ray Absorption Spectroscopy

Synchrotron techniques

It is possible to classify, grosso modo, the techniques exploited at synchrotron facilities in four groups:

- **Diffraction and scattering**, including Small- and Wide Angle X-ray Scattering (SAXS/WAXS), powder and monocrystal diffraction, passing trough imperfect stressed crystals, protein crystallography, X-ray reflectivity
- **X-ray Spectroscopies,** for analysis of chemical composition, electronic structure, and local structure around a selected site. It includes the X-ray emission and absorption spectroscopies, enclosing a multitude of techniques (Extended X-ray Absorption File Structure (EXAFS), X-ray Absorption Near Edge Structure (XANES) …). Other techniques include X-ray photon correlation spectroscopy (XPCS), inelastic and Raman scattering, Compton scattering, etc.
- **Imaging**, including also tomography, with many detection techniques (absorption, phase-contrast, diffraction enhanced, fluorescence, diffraction, holotomography, etc.)
- **Others**: Medical therapy, Polarimetry …

In the following I will restrict the discussion to two basic and important techniques: Powder diffraction and absorption spectroscopy. Only general concepts are given, for more information I address the reader to the web sites of the synchrotron facilities and to the vast literature.

X-ray Diffraction using 2D detectors

A typical experimental setup for X-ray diffraction using an image detector (a CCD camera, image plate, film) is in Figure 3.

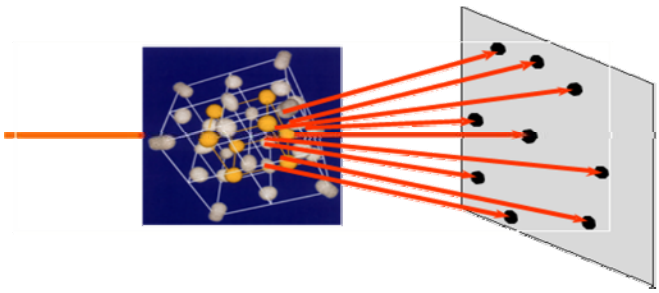

Figure 3 Normal transmission setup of an X-ray scattering experiment for monocrystals.

The primary (monochromatized and possibly focused or collimated) beam impinges the sample, which is usually small. In the case of sample powders they are usually placed in a glass capillary). The sample diffracts the X-rays following the laws of crystallography. For the case of a monocrystal, the excited reflections are those points in the reciprocal lattice that are intercepted by the Ewald sphere. These intersection points give the directions of the diffracted beams that are intercepted and registered by the detector (Figure 4).

An example of a diffraction pattern for a monocrystal is in Figure 5. If the sample consists of a large collection of randomly orientated single crystals (i.e., a powder), the diffracted beams are seen to lie on the surface of several cones with axis along the incident beam. The cones emerge forwards and backwards. The detector, if

placed perpendicular to the beam) will register concentric circles, as illustrated in Figure 6.

In Figure 7 it is shown a diffraction pattern from alumina powder, also measured at ESRF (ID18F beamline). It can be observed some inhomogeneities in the rings, which reveal the grain structure of the powder. Even if this is a very fine powder, the small x-ray beam (dimension of some tens of microns) reveal discrete grains.

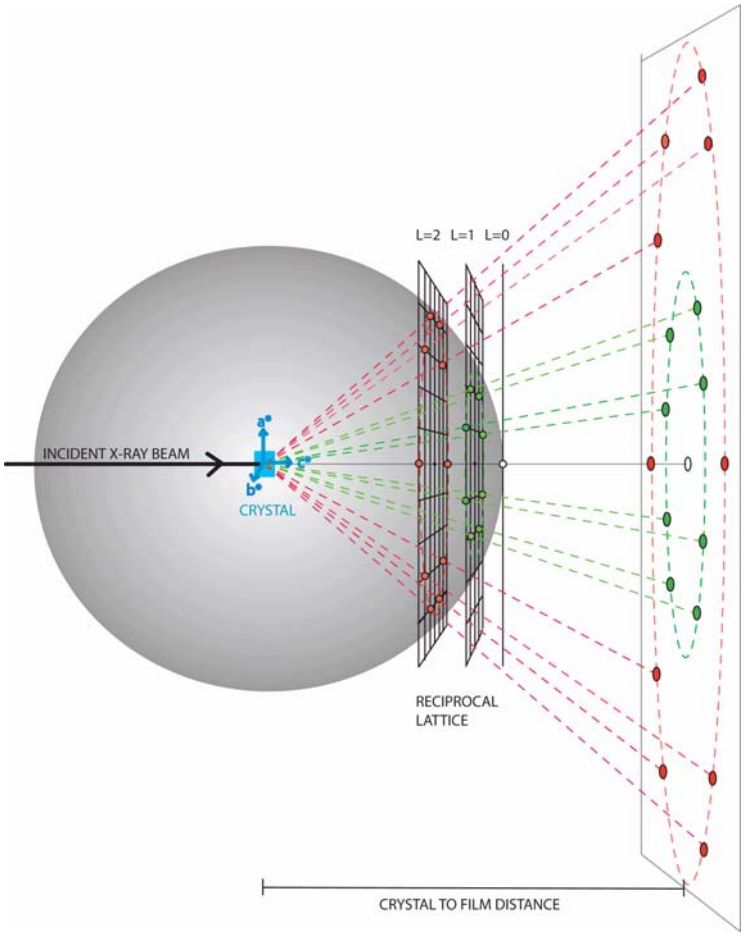

Figure 4. In the case of a monocrystal, the points of the reciprocal lattice that intersect the Ewald sphere originate the diffracted beams that are recorded by the detector. (Figure from http://www.doe-mbi.ucla.edu/~sawaya/m230d/Data/ewald_mikey2.gif).

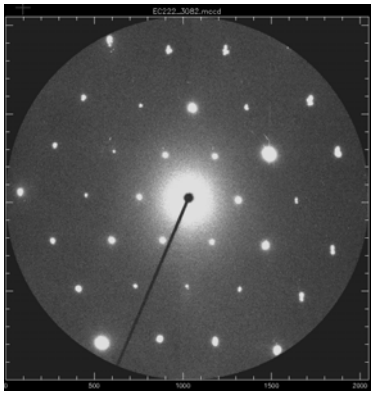

Figure 5: An example of the Laue diagram showing the pattern of X-rays diffracted by a single crystal of mica. The small intense spots lined crosswise are the so-called Laue spots. This image has been recorded at the ID18F beamline, showing the typical hexagonal pattern. For more information on mica structure see, for example, Nespolo et al. (1997).

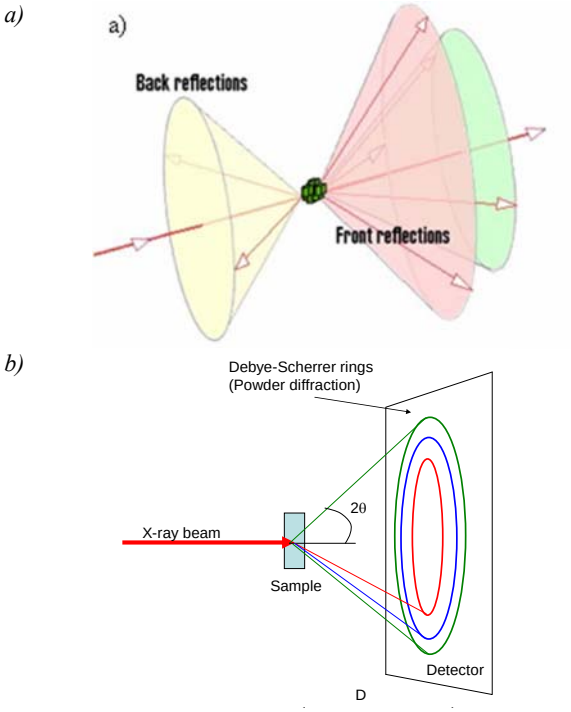

Figure 6. Origination of concentric rings (the Debye-Scherrer rings) in powder diffraction (Figure 6a from http://www.matter.org.uk/diffraction/x-ray/powder_method.htm).

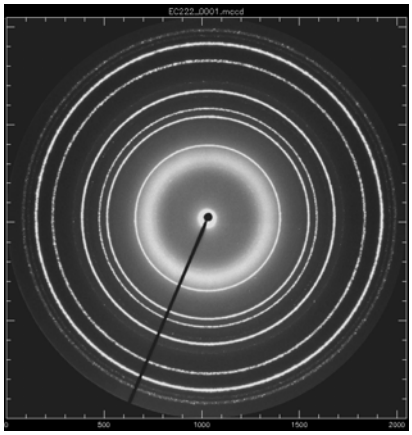

Figure 7 Diffraction pattern for alumina recorded at EXRF (ID18F beamline).

The two cases presented (monocrystal and powder) are the extreme case of many possible variations that can be found in the samples. The beauty of the use of image detectors is the possibility to *see* many aspects related to the sample by simply inspecting the diffraction images. The position of the rings are related to the crystal structure. The granularity indicate the crystal coarseness. The strain in materials (i.e., the deformation due to the application of a load to the sample) is characterized by a small loss of the cylindrical symmetry of the rings, showing deformations along particular directions. Diffraction by long structures aligned along a given direction (fiber diffraction) will produce some typical patterns with symmetry along an equatorial line (Figure 8). In fiber diffraction the scattering pattern does not change, as the sample is rotated about the fiber axis.

Fine powders will produce Debye-Scherrer rings very homogeneous, whereas that samples containing inhomogeneous powder will produce images with strong spots meaning that large grains reflect as “monochrystals”. Of course the “fines” of the powder is related to the size of the beam, and a very small beams will reveal more details about the particular grains. For a non-ideal powders, the diffraction patterns give information on the crystallographic structure, the crystallite size (grain size), and preferred orientation in polycrystalline or powdered solid samples.

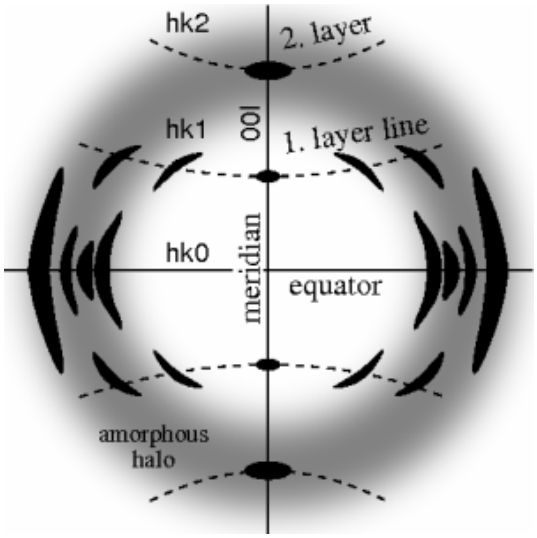

Figure 8 Ideal fiber diffraction pattern of a semi-crystalline material with amorphous halo and reflexions on layer lines. High intensity is represented by dark color. The fiber axis is vertical. (from http://en.wikipedia.org/wiki/Fiber_diffraction)

The standard powder diffractogram can be retrieved from the scattering image by performing an azimuthal integration around a center that is the intersection of the direct beam with the detector plane. That means that we eliminate the angular dependency and display only the intensity versus a radial distance from the center of the rings. The radial distance can be easily converted in scattering angle (2θ) once we know the geometry of the experiment.

The use of image detectors has some advantages with respect to the diffractometers that usually record θ-2θ scans. The access to the image gives a lot of information of the sample heterogeneity. Also, the complete diffractogram is obtained in one single shot, and this opens the possibility of performing time-dependent, temperature-dependent, or in general study the evolution of the diffractogram as a function of one selected parameter. However, the data analysis is more elaborated and care must be taken in order to produce diffractograms that could be compared to those measured in conventional or synchrotron diffractometers.

The steps that are usually followed to obtain integrated diffractograms from diffraction images are the following:

- Correction of the image distortion caused by the optical aberrations of the detector
- Substraction of the dark field image
- Normalization with the flat field image
- Calibration of the image using a well characterized reference
- Masking the image to remove pixels not illuminated correctly (beam stopper, etc.) or saturated (spots due to monocrystals).
- Integration of the image, supposing the experimental geometrical characteristics did not change (sample and detector positions)

The analysis of diffraction and scattering data using 2-dimensional detectors requires the precise knowledge of the intersection point beam-detector and the distance from sample to detector, assuming that the beam direction is orthogonal to the detector plane. In most experiments, these parameters are obtained from a non-linear minimization procedure using an image from a powder reference sample. The procedure usually allows for small corrections in the angle beam-detector, which is 90º only for an ideal case. These parameters are used for integrating the images. An azimuth integration of the image will produce the intensity versus 2θ diffractogram that will be further analyzed by different methods. The quality of the diffractogram depends very much on the accuracy of these geometrical parameters.

If the detector is not completely perpendicular to the beam direction, the rings appear as ellipses instead of circles. Here, for retrieving the integrated diffractogram, it is necessary to know with high precision how the detector is positioned with respect to the beam. This can be done with the code XOP/Xplot2D.

For high-resolution powder diffraction experiments, it is in general preferred to used the standard θ -2 θ scan methods, rather than using image detectors, because they are more precise in terms of angle and position, thus obtaining in general a better and

more controllable resolution in the diffractogram, and instrumental functions of better quality.

Absorption spectroscopies

The X-rays are scattered by with matter following different physical interactions (Figure 9).

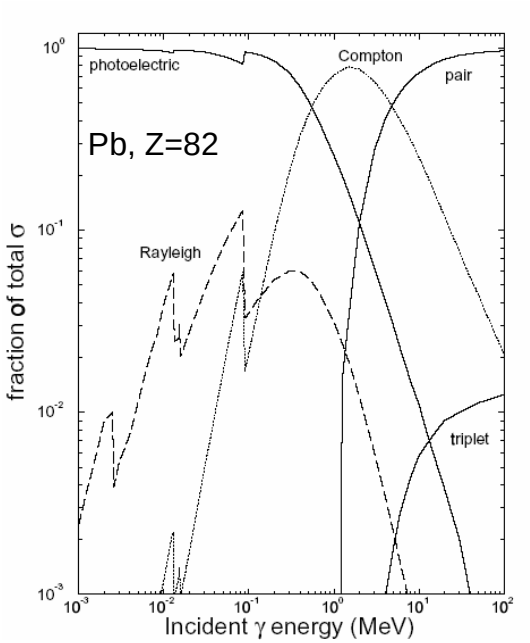

Figure 9 Photon interactions with matter (fractional cross sections for led) as a function of the photon energy.

We recall that diffraction is originated by the interference of x-rays elastically scattered (i.e., Rayleigh) by the atoms in the crystal. The bind energies of the inner electrons in atoms go from a few tens of eV to several keV. The photoelectric absorption is the dominant interaction for photons with energies in this interval (Figure 10).

The X-ray photoelectric absorption coefficient for an atom μ (see Figure 10) is proportional to the probability that this atom absorb a photon (cross section). It

decreases monotonically with photon energy, but shows discontinuities (jumps) at the absorption edges, i.e., when the photon has enough energy for removing an electron from its bounded state (orbit around the nucleus).

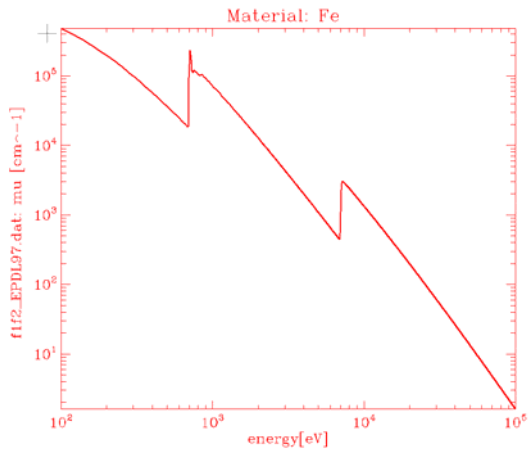

Figure 10. Photoelectric absorption versus photon energy for iron. It can be shown the transitions corresponding to the excitation of the inner electrons (absorption edges: K ($1s_{1/2}$) at 7112 eV, L_1 ($2s_{1/2}$) at 846.1 eV, L_2 ($2p_{1/2}$) at 721.1 eV, L_3 ($2p_{3/2}$) at 708.1 eV).

Each element on the periodic table has a set of unique absorption edges corresponding to different binding energies of its electrons. However, a fundamental fact is that the form of the absorption edge is different depending on local neighborhood of the absorbing atom. For instance, the K-edge of Fe is very different in metal iron (each iron atom is surrounded by eight iron atoms) than in Fe oxides, where an iron atom is surrounded by six oxygen atoms in octahedral configuration (See Figure 11). This variability of the absorption edges depending on the atomic environment is the origin of the X-ray Absorption Spectroscopies (XAS). The possibility to select a given atomic specie by tuning the x-ray energy gives the XAS element selectivity.

The interest of absorption spectroscopy for condensed matter comes from the complex structure of the absorption edge. This XAFS (X-ray Absorption Fine Structure) contains information of the electronic structure and arrangement of the neighborhood atoms around the selected excited atom. By analyzing the XAFS,

information can be acquired on the local structure and on the unoccupied electronic states.

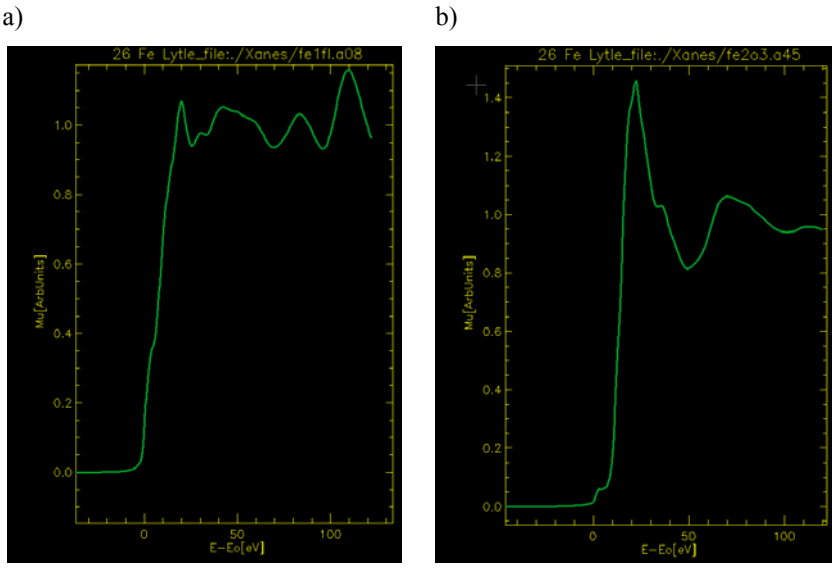

Figure 11 K-edge X-ray absorption for iron, a) in a metallic iron sample (iron metal packs in a body-centered cubic structure, therefore an Fe atom is surrounded by eight Fe atoms), and b) in an iron oxide Fe_2O_3 (the Fe is in a center of an octahedron with six oxygens). Data comes from the Farrel Lytle database (http://xafsdb.iit.edu/database/data/Farrel_Lytle_data/), available in XOP/XAID.

In a XAS experiment the intensity of the x-ray beam after interacting with the sample is monitored versus x-ray energy. The most basic setup works in transmission. The energy of the X-ray is scanned by the monochromator, and there are two intensity detectors (usually ionization chambers) one before the sample, that measures the incident intensity $I_0(E)$ and the other after the sample that measures the transmitter intensity $I(E)$. The dependence of the absorption coefficient on the photon energy □□ is given by is the Lambert-Beer law:

$$I(E) = I_0(E)e^{-\mu(E)x} \qquad (0.1)$$

being x the sample thickness. If the readings of the detector (ionization chambers) are proportional to the intensity, the sample thickness can be optimized to give $I / I_0 \approx\sim 0.1$ (Parrat et al. 1957)

The absorption coefficient is obtained by taking the log ratio of the incident x-ray intensity to the transmitted x-ray intensity.

$$\mu = -\frac{\ln(I / I_0)}{x} \qquad (0.2)$$

The photon energy is scanned around the absorption edge of an element included in the sample. Typical scans go from about 100 eV below the edge to about 1000 above it for EXAFS (about 100-220 eV for XANES) with steps of about 1 eV for EXAFS (0.1-0.2 for XANES).

The X-ray absorption spectra show a steep rise at the core-level binding energy of X-ray-absorbing atoms and attenuates gradually with the X-ray energy. The XAS spectra are usually divided in three energy regions: 1) the edge region, 2) the X-ray Absorption Near Edge Structure (XANES); 3) the extended X-ray absorption fine structure (EXAFS). The absorption peaks at the absorption edges very close to the edge (in a range of about 5 eV) are due to electronic transitions to first unoccupied molecular levels above the chemical potential. The XANES part extends over a range of about 100 eV above the absorption edge. The information on the geometry of the local structure is provided by the analysis of the *multiple scattering* peaks in the XANES spectra. The oscillatory structure extending for hundreds of electron volts past the edges is called EXAFS, and contains information on the number and type of atoms in the firsts coordination shells around the photoabsorber.

XAFS spectra are most often collected at synchrotrons. Because X-rays are highly penetrating, XAS samples can be gases, liquids or solids. And because of the brilliance of the synchrotron X-ray sources, the concentration of the absorbing element can be as low as a few parts per million in weight (ppm).

In addition to standard XAFS beamlines, new experimental stations equipped with fast scanning options (Quick-EXAFS) or energy-dispersive XAFS beamlines provide time resolution in the sub-second range. XAS is used in a growing number of time-resolved experiments, kinetic studies and mapping application (also called hyperspectral, as each mapped point is a complete XAS spectrum). Since the number of spectra measured during one multiple-scan experiment can amount to several hundred, such investigations demand data-reduction software prepared to cope with large data sets. For the data analysis purposes, XOP includes a package called XAID, which includes numeric tools for spectra normalization, EXAFS extraction, together with a versatile and efficient graphical user interface.

EXAFS data analysis

The EXAFS region results from the interference in the *single scattering* process of the photoelectron scattered by surrounding atoms. It gives information on:

- Radial distances
- Coordination numbers
- Types or nature of neighborhoods

Basically every XAFS data analysis begins with the visualization of experimental spectra and basic reduction steps such as energy calibration, and background subtraction.

The goal of the data analysis is to reveal the characteristic parts of the fine structure characteristic condensed matter, i.e., eliminating the contribution of the atomic absorption. The atomic absorption, instrumental background, and absorption from other edges are usually eliminated by performing a pre-edge and post-edge fits. The pre-edge is usually fitted to a straight line that is removed from the spectrum. A more precise fit is the so-called Victoreen polynomial, but results are very similar for typical pre-edge lengths (<100 eV). The experimental absorption edge is usually placed at the inflexion point (at the maximum of the derivative). The post-edge for XANES may be fitted to a constant (i.e., *jump* value *J*) or to a polynomial (first or second degree).

For EXAFS, the spectrum abscissas are converted from energy to wavenumber of the photoelectron using the formula:

$$\left|\vec{k}\right| = \frac{2\pi}{\lambda} = \sqrt{\frac{8\pi^2 m}{h^2}(E - E_0)} \qquad (0.3)$$

being m the electron mass, E_0 the experimental edge energy and h the Planck constant.

For EXAFS, the post-edge is fitted, usually in k space, with several polynomial splines, that are matched among them (continuity or the functions and continuity of the first derivative). This operation is very delicate, as it is necessary that the splines will pass among the oscillations of the fine structure. Then, the x-ray absorption signal is normalized to unit step height to obtain the experimental EXAFS signal $\chi(k)$. Several algorithms are available: i) experimental, ii) constant, and iii) Lengeler-Eisenberg:

$$\chi_{exp}(E) = \frac{\mu(E) - \mu_{spline}(E)}{\mu_{spline}(E) - \mu_{pre-edge}(E)} \qquad (0.4)$$

$$\chi_{cte}(E) = \frac{\mu(E) - \mu_{spline}(E)}{J} \qquad (0.5)$$

$$\chi_{L.E.}(E) = \frac{\mu(E) - \mu_{spline}(E)}{J\left(1 - \frac{8}{3}\frac{E - E_0}{E_0}\right)} \qquad (0.6)$$

If the ejected photoelectron is taken to have a wave-like nature and the surrounding atoms are described as point scatterers, it is possible to imagine the backscattered electron waves interfering with the forward-propagating waves. The resulting interference pattern shows up as a modulation of the measured absorption coefficient, thereby causing the oscillation in the EXAFS spectra. A simplified plane-

wave single-scattering theory is used for interpretation of EXAFS spectra. In the single-scattering formalism, the EXAFS signal can be written as (Teo 1986):

$$\chi(k) = \sum_j N_j \; S_j(k) \; F_j(k) \; e^{-2\sigma_j^2/k^2} \; e^{-2r_j/\Gamma_j(k)} \; \frac{\sin(2kr_j + \phi_{i.j}(k))}{kr_j^2} \qquad (0.7)$$

where F is the backscattering amplitude for each of the N_j neighboring atoms of the jth type with a Debye-Waller factor σ_j (to account for thermal vibrations and static disorder), $\emptyset$ is the total phase shift experienced by the ejected electrons (photoelectrons). The second exponential term includes the inelastic losses of the scattering process, with Γ the electron mean free path. S is an amplitude reduction factor (S~0.8) due to many body effects.

From this equation, the EXAFS signal is made by a sum of sinusoidal signals, one per surrounding coordination shell (indexed by j) multiplied by the coordination number N_j and amplitude factor F, among other damping terms. The frequency of the sinusoidal components is determined by the distance to the photoabsorber r_j, and is affected by a phase shift $\emptyset$.

A full analysis of the EXAFS spectrum consists in fitting the experimental EXAFS signal $\chi(k)$ to Eq. (0.7) and minimizing a number of free parameters (usually N, r, σ, Γ). The curve-fitting analysis needs the scattering amplitude F and phsase shifts $\emptyset$ that can be obtained from three different methods:

- From the analysis of the EXAFS data of the reference materials, supposing the "phase-transfer" hypothesis. Note that the amplitude function F depends only on the type of backscatters, but the phase function $\emptyset$ contains contributions from both the absorber and backscatterer.
- From tabulated data, like the McKale et al (2002) XAFS phase shift tables available in XOP/XAID
- From ab-initio calculations, using modern methods (like FEFF, GNXAS).

Ab-initio calculations are the most precise method up to date, which may also account for curved-wave corrections and multiple-scattering effects that sometimes can not be neglected.

The Fourier transform is critical to EXAFS analysis. In fact, applying the Fourier transform to (0.7) is easy to see that it will present peaks at the coordination shell distances r_j (not exactly, as there is a phase shift), and the intensity of the peaks is proportional to the coordination number N_j. Thus the Fourier transformed EXAFS can be used to isolate and identify different coordination spheres around the absorbing atom. The Fourier Transform is a complex function (it has a real and imaginary part). It is common to display only the modulus as shown in Figure 12

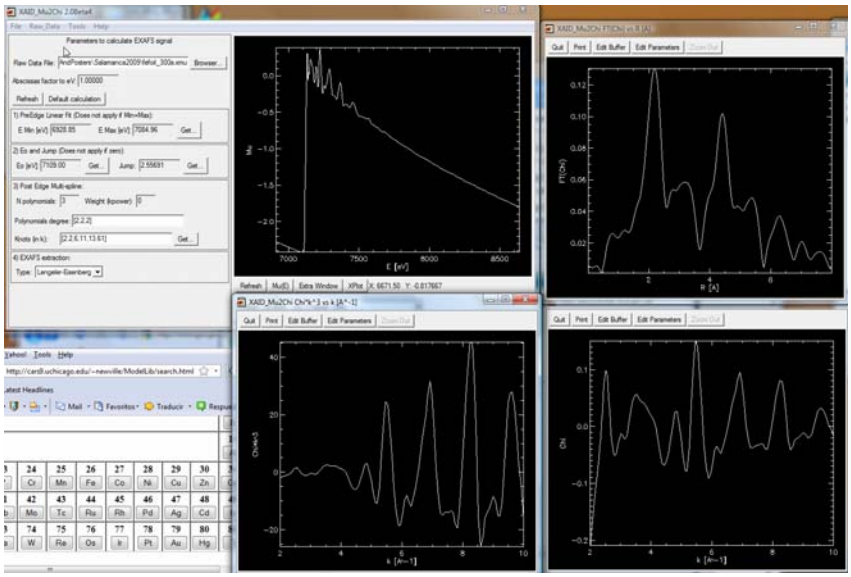

Figure 12. EXAFS data analysis using the XOP/XAID application. The experimental EXAFS corresponds to a Fe foil and has been obtained from the database: http://cars9.uchicago.edu/~newville/ModelLib/search.html (file: fefoil_300a.xmu).

The absorption $\mu(E)$ can be measured in transmission beamlines, as already mentioned. However, the the fluorescence or electron emission signals are also modulated with XAS, and can be used as recording signals. When the matrix is very

absorbent (it contains heavy elements), or the photoabsorbing element is very dilute it is preferred the use of fluorescence detector. Electron yield detector is well suitable at low energies.

Information from XANES and data analysis

The pre-edge, edge structure and the absorption to a few hundeds of eV above the edge (all of them grouped under the same denomination, XANES) contain a rich information on:

- The chemical oxidation state
- The bonding states (covalency)
- The site symmetry and local environment (distances and angles)
- The coordination structure

The pre-edge and edge structures depend on the electronic structure. The pre-edge features are caused by electronic transitions to empty bound states. The shape and width of the main edge is originated by the transitions of the inner electrons to a wide band of allowed states beyond the Fermi level. The position of the edge is affected by the oxidation state (chemical shift). The photelectron scattering amplitude in the low energy range (5-200 eV) of the phoelectron kinetic energy become much larger than for the high energy range (EXAFS) so that multiple scattering events become dominant in the XANES (sometimes called NEXAFS) spectra. Therefore, there is not only information on the coordination shells, as in EXAFS, but also on the nature of the individual atoms, and their positions (distances and angles). At these low energies, the photoelectron is affected by the outer electrons energy state, therefore it is possible to obtain information on oxidation states, covalency, etc. The drawback of this rich information contained in the XANES part of the X-ray absorption spectra is that it is more difficult to obtain quantitative data than in EXAFS, where the information is less rich but the analysis is simpler. Therefore, most of the XANES analysis is mostly qualitative, aiming to compare different spectra and obtain some conclusions from their differences.

A full quantitative data analysis of the XANES involves the proposition of a structural model and compute the theoretical XANES using ab-initio computer

programs, such as CONTINUUM, MSXAS, MXAN or FEFF. Then the model may be refined manually, or even fitted automatically as in MXAN (Benfatto and Della Longa (2001).

XANES contains very rich information, and in many times a qualitative analysis of XANES is enough to extract part of this information comparing XANES spectra from different samples or from the same sample under different conditions. These fingerprint analyses involve one or several of the following points:

- Analysis of the pre-edge peaks. These pre-peaks may be fitted to pseudo-Voigt functions for studying their evolution and changes. A pseudo-Voigt is a linear combination of a Gaussian, that accounts for the instrumental functions, and a Lorentzian, that accounts for an atomic transition.
- Analysis of the edge shift: study the change of the position of the experimental edge (inflexion point).
- Analysis of the main edge. A white line (strong peak after the edge) is usually observed in the K-edge spectra of transition-metals compounds and L-edge of rare-earth compounds. The XANES may be fitted to an arc-tangent function (to account for the edge-step) plus one or several Lorentzians or pseudo-Voigt functions (to account for the atomic transitions).
- Analysis of the evolution of the different structures shown in the XANES region, and relate them to multiple scattering effects.
- Analysis of the XANES of a mixture as a weighted fit (linear combination) of the XANES of the individual components (ingredients).
- Statistical analysis (such as Principal Component Analysis)of a large number of XANES spectra.

XOP

XOP (X-ray oriented programs) is a user-friendly computer environment for performing calculations of interest to the synchrotron radiation community. It is freely available. It provides codes for: i) modeling x-ray sources (e.g., synchrotron sources, such as undulators and wigglers), ii) calculating characteristics of optical elements

(mirrors, filters, crystals, multilayers, etc.), and iii) multipurpose data visualization and analyses. Only the third point will be discussed in this paper.

For downloading and installing XOP consult the web pages: http://www.esrf.eu/UsersAndScience/Experiments/TBS/SciSoft/xop2.3/ Screenshot of the main XOP window are in Figure 13

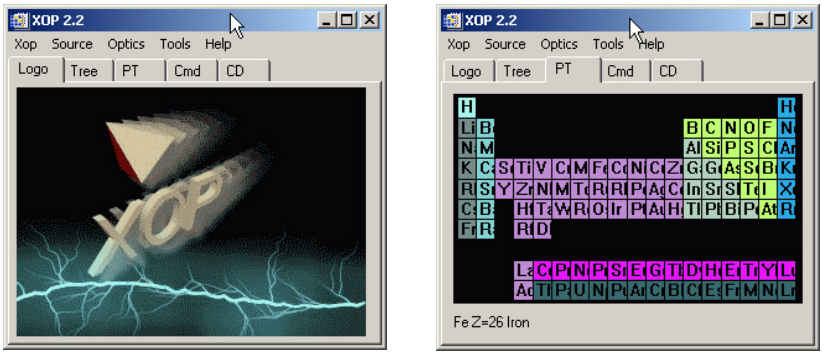

Figure 13. Main window of XOP. a) Stating window with logo, b) periodic table.

XOP for Data Analysis

A number of generic tools are available in XOP to visualize and analyze numerical data. There are interconnections among them (Figure 14). The most important tools are:

- XPLOT is an XY-data analysis and visualization generic tool with several options to make overplotting, fitting, high-quality customized graphics, etc.
- EXODUS can be used to import multiple sets of data from either SPEC files or multiple ASCII files, and permits to interpolate, average, operate and display these sets of data in a very flexible way.
- XPLOT2D is an imaging tool targeted for scientific image display. It can import data in a variety of formats (ESRF data format, MAR CCD, TIFF, etc) and allows different selection of region of interest (ROI), masking data, radial and azimuthal integrations, etc.

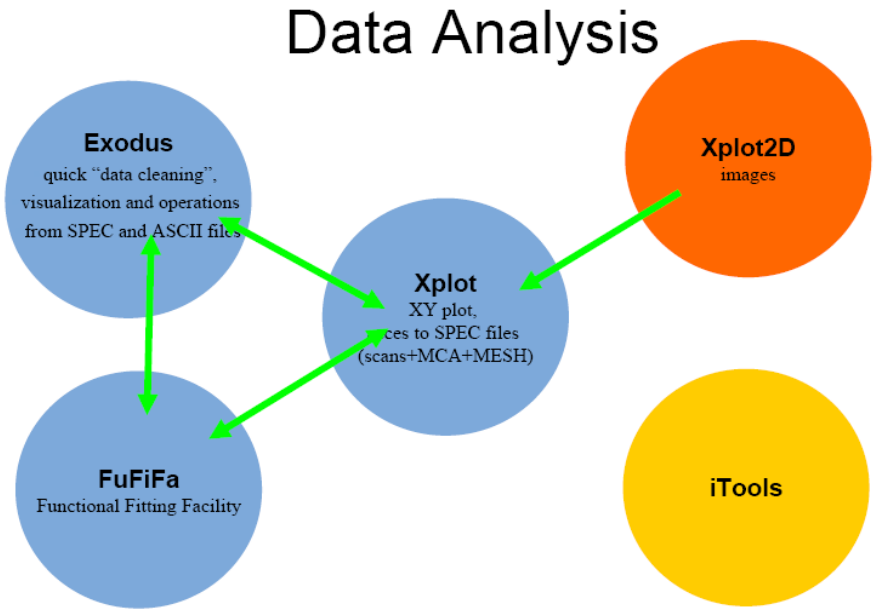

Figure 14. Schematic connectivity among XOP's data analysis tools.

Regarding some particular synchrotron radiation techniques, some tools are available:

- DABAX (DAtaBAse for X-ray applications) is a front-end interface to the data files containing tables for photon-atom interactions. It allows one to i) inspect data files, ii) download new data files to your local XOP installation, iii) consult file documentation, iv) view numerical data and, v) plot the data files, and vi) start the application programs that calculate interactions of x-ray beams with optical elements or components.
- For x-ray crystallography:

  XPLOT2D can be applied to the analysis of image plates with X-ray diffraction data. It permits the calibration of some detector parameters (position of the center of the beam and distance sample-

detector), and it allows overplotting on top of the recorded image the Debye-Scherrer rings of a known material (see Figure 15).

XPOWDER permits to simulate powder diffraction patterns starting from a known crystal structure.

- For x-ray absorption spectroscopies: the XAID *extension*[1] *of XOP* included data analysis tools for x-ray absorption spectroscopy (background substraction, EXAFS signal extraction, Fourier Transform, Fourier Filtering) as well as a database of experimental XAFS spectra.

a) b)

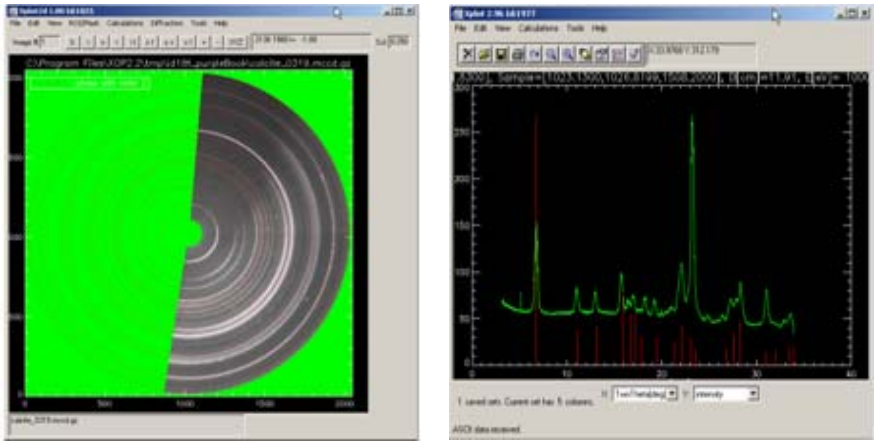

Figure 15. a) Example of a powder diffraction image (left) of an archaeological pigment (Maya blue) with overplotted rings corresponding to a given reference (palygorskite). b) azimuthal integration of the image to obtain the diffractogram, and overplotted peak position of the reference.

The XOP interface is written in the Interactive Data Language IDL (http://www.ittvis.com/), and it runs on the Linux Windows and MacOS operating systems. It has been built with an IDL license embedded and is available free of charge.

[1] *An XOP "extension" is a software package that is not part of the XOP standard distribution but can be optionally installed and run under the XOP interface. The complete list of extensions, download instructions, tutorial, and installation procedures is available from the XOP web page.*

In addition to the on-line applications available from the XOP interface, some programming capabilities are allowed using the "XOP Macros". These are short programs in IDL-like syntax that permit easily to perform customized repetitive tasks. An example of a macro used to create Figure 1 is found in Appendix 1. Users interested in developing their own programs profiting from the available XOP tools can use XOP in "development mode", with full access to the XOP library of IDL routines, but in this case the user must have access to a valid IDL license.

Examples

The data analysis tutorial document (available in the XOP web site) describe some commented examples to:

- Generic data analysis of spectral data: import data, manipulation, visualization, generic analysis (background removal, etc.), fitting and modeling
- Analysis of diffraction image (powder diffraction): visualization, calibration, integration (see also the on-line documentation of Xplot2D).
- XAFS data analysis: XANES normalization and fingerpring analyses. EXAFS extraction, Fourier transform (see also the on-line documentation of XAID).

Appendix 1

XOP macro to generate Figure1

```
x=findgen(101)-50 & z=exp(-x^2/200.)#exp(-x^2/200.)
isurface,z
xplot2d,z,scale=5
window,/free,xsize=500,ysize=500
surface,z
window,/free,xsize=500,ysize=500
contour,z
```

References

Benfatto, M. and S. Della Longa, (2001). Journal of Synchrotron Radiation, 8(4): p. 1087-1094.

McKale, A.G., B.W. Veal, A.P. Paulikas, S.K. Chan, and G.S. Knapp, (2002) Journal of the American Chemical Society,. 110(12): 3763-3768.

Nespolo, M., H. Takeda, G. Ferraris, and T. Kogure, Mineralogical Journal, (1997). 19(4): 173-186.

Parratt, L.G., C.F. Hempstead, and E.L. Jossem,.(1957) Physical Review, 105(4): 1228.

Teo, B.K., (1986) ,*EXAFS: Basic Principles and Data Analysis*. Berlin: Springer-Verlag. 349.