

To be published in the EPDIC7 proceedings.

[7th European Powder Diffraction Conference](#), Barcelona, 20-23 May 2000.

Structure Determination by Powder Diffractometry : Internet Course

Armel Le Bail¹, Yvon Laligant¹ and Alain Jouanneaux²

¹ Université du Maine, Laboratoire des Fluorures, CNRS ESA 6010, Avenue O. Messiaen,
72085 Le Mans Cedex 9, France

² Université du Maine, Laboratoire de Physique de l'Etat Condensé, CNRS ESA 6087,
Avenue O. Messiaen, 72085 Le Mans Cedex 9, France

Keywords: teaching, structure determination, powder diffraction, Internet, distance learning

Abstract Details on a distance learning course on the Internet about structure determination by powder diffractometry are given.

Introduction New specialties, by definition, cannot be represented by top notch teachers at work everywhere. Internet offers new oportunities for efficient distance education. Structure determination by powder diffractometry (SDPD) is a recent topic which could not list more than 60 solved experimental cases ten years ago. Today, real studies are close to 500, and techniques continue to evolve fast. Few tutorial exist explaining how experts deal now with powder samples from the synthesis to the Rietveld refinement, passing through steps such as identification, indexing, structure factors extraction, and use of Patterson, direct or molecule location methods. A widely visited Web site offers pages entitled "Strategies in Structure Determination from Powder Data" since 1995 [1], including many exercises with solutions. On that basis, a 100% distance learning course has been built [2], adding more documents and pedagogy. The course is asynchronous, registration is opened all the year, and each student makes progress at his own speed. The three teachers totalize 10% of the published SDPDs since 1988. They propose here a new series of exercises based on unpublished real cases, distributed together with a Web-based documentation in 10 course sessions detailed in this paper. The continuous assessment of the students progress is based on their solutions to these exercises. The minimum course duration is of 12 weeks (expandable to one year, depending on the student time availability), including the final examination. A University of Maine diploma is delivered in case of success to both the continuous assessment (counting for 50%) and the final examination which consists in realizing a final SDPD in two weeks. A key to the success is fast interactivity with the teachers, by e-mail. Interactivity with the community of experts in the field is possible through the SDPD [3] and the Rietveld [4] mailing lists, having more than 260 and 300 subscribers, respectively. The course tries to honestly show the diversity of available programs and can almost be realized without commercial software. However, the access to some databases, and programs able to search in them, is needed at least for (failing in) sample identification.

Course program Books and publications contain the mathematical and geometrical basis of crystallography. Tutorials on the Web propose more hints and tips that allow to learn how to succeed. There is not enough time during traditional courses in faculties for really showing how a powder diffractionist realizes a complete SDPD, because this may be a job of several days, if not several weeks, and may fail. The SDPD Internet Course is thus organized for pointing toward

the pertinent books and references. Links to documents as well as to the broadest range of software are also given. The 10 sessions are organized linearly as detailed below.

Session 1 That first part is in free access at the SDPD Internet Course Web site [5], as a demonstration. It is devoted to powder pattern recording conditions, phase recognition, use of specific databases. The material includes parts of tutorials, conferences on line, chapters of books to read, exercises with solution showing how modern search-match software can help to identify phases by using the ICDD PDF-2 database. Unfortunately, no free access search-match software can be proposed here (the student is informed of that before subscribing). The software to download and test at session 1 concern mainly conversion programs. It is essential to be able to convert powder data to any format that can be needed at any time. Some available programs which can do that are also able to perform peak fitting, or more, but can open and save data files in various formats. Being a programmer would help to build simple softwares like *DAT2RIT* (Fortran, open source) or to modify them for special purposes. Names of such conversion-able programs follow : *CONVERT*, *POWDER*, *WINFIT*, *CONVX*, *POWDERX*, *AXRLEC*... We have no place here to give all the software references, most of them will be located by using their name as a keyword in the main Internet search engines (Altavista, Google, Yahoo...). The student cannot expect to obtain perfect help for all programs listed along the SDPD Internet Course, because the teachers cannot be familiar with all of them. Asking for help for a given program should be addressed to its author, after a careful reading of the manual, and examination of the above-cited mailing lists archives. Exercises consist in trying to identify the phases present in 6 samples of which the powder patterns are provided. What we expect from the student is not always an exact result, though this would be possible sometimes, but an estimation depending on the PDF-2 database completeness.

Session 2 The subject is "Indexing powder patterns, part 1." Documents and conferences online are provided together with a series of data which the student is invited to play with before trying to solve the session exercises. The examples are $\text{Na}_2\text{C}_2\text{O}_4$, $[\text{Pd}(\text{NH}_3)_4]\text{Cr}_2\text{O}_7$, $\alpha\text{-AlF}_3$, $\beta\text{-BaAlF}_5$ and cimetidine ($\text{C}_{10}\text{H}_{16}\text{N}_6\text{S}$), as well as 2 samples from the SDPD Round Robin [6]. Indexing supposes peak position extraction, and most diffractometers are delivered with a commercial software. But the student may want to try some others. Very frequently, hyperlinks in the SDPD Internet Course HTML documents will connect the student to the CCP14 Web pages [7], containing the biggest useful amount of information about powder diffraction and small molecule structure determination on the Internet. Links to programs *POWDER*, *WINFIT*, *POWDERX*... are provided. Indexing software may be found in several places. The most up to date versions should always be obtained from the authors themselves. However, the latest versions may also be found at some repository places. *TREOR*, *ITO* and *DICVOL* continue to be distributed together with their Fortran source code. The full list of indexing programs that were used in successful SDPD is found in the SDPD-Database [8]. Other programs that the student may need at this time are those who can generate the whole list of reflections for a given cell (*HKLGEM*, *ERADIS*...). For instance, if you choose to add a standard compound like NAC ($\text{Na}_2\text{Ca}_3\text{Al}_2\text{F}_{14}$) to your sample, in order to calibrate, then you need the full list of the NAC reflections. In fact, many programs can deliver such *hkl* list of reflections together with angular positions (for instance some Rietveld codes), but hyperlinks to them will be given a few sessions later. Unit cell refinement (*LAPOD*, *CELREF*, *ERACEL*...) may be an important step in the process, both for calibration when a standard compound is mixed with the unknown, or for refining the cell parameters as proposed by the indexation results (though a more sophisticated way exists now, which will be seen in a future session during the course). Some software are free for non-commercial use but cannot be obtained without contacting their authors. This may take time, so that the students are informed here that they should obtain now programs for future session (*EXPO*, *SIR*, *SHELX*, *WINGX*, *DIRDIF*...). The exercise of session 2 needs an exact result, probably to be located in a long list of wrong propositions. The compound is a sodium

chromium fluoride. Two powder patterns are given, one including a standard (*NAC*). The student has to propose a zero point for both patterns and should produce the *TREOR*, *ITO* and *DICVOL* indexations, if possible. At this stage, teachers do not ask for more (the student has not to propose any space group).

Session 3 The session title is "Indexing powder patterns, part 2." It introduces the student to the most complete package, *CRYSFIRE* from Robin Shirley, grouping 8 indexing programs. A synchrotron pattern is provided, corresponding to a pure phase (no mixed standard), and it is time to see if the student becomes an expert in indexing. Moreover, a second synchrotron powder pattern, which was disclosed at the SDPD Mailing List [3], is also provided, still waiting for a convincing solution from the world best experts, including us ! Checking of multiple cell propositions by *CRYSFIRE* can be done by using the new *CHEKCELL* program.

Session 4 After a minimum of 2 weeks on indexing problems, the student goes further in a new subject : "Advanced methods for extracting structure factors, part 1." A good introduction is in the online conference "The Practice of "|Fobs|" Extraction from Powder Diffraction Data" [9]. Powder pattern decomposition is treated in the Rietveld method book (individual profile fitting or Pawley method in chapter 14). Chapter 15 is devoted to *ab initio* structure solution and cites some works where the Le Bail method was first used. Another book is still waited since the SDPD-95 workshop, Oxford. The students have already extracted intensities in previous sessions, because intensities are a by-product of peak position estimation. However, transforming those intensities in amplitude structure factors needs the cell knowledge and the attribution of *hkl* Miller indices to each extracted amplitude. The list of programs used in SDPD for this purpose is quite long. The Le Bail method is more distributed than the Pawley method, embedded in commercial programs (*DASH* or *POWDERSOLVE* for instance). However the Cambridge Fortran resources may provide the Pawley program Fortran source code. Many programs below are Rietveld programs including the Le Bail fit algorithm which consists in iterating the Rietveld decomposition formula, starting from a set of arbitrarily equal "|F|", instead of the "|Fcalc|" obtained from the starting structure model. The full list of programs used for this purpose of pattern decomposition and which were used in SDPD real cases, is online in the SDPD-Database [8]. Many of the following programs, certainly able to do the job, are available directly on the Internet : *FULLPROF*, *GSAS*, *WINMPROF*, *EXPO*, *EXTRAC*, *LHPM/RIETICA*, *XND*, *ARIT*, *ALLHKL*, *SIMPRO*, *WPPF*... At this stage, other programs which may be useful can display a powder pattern and will allow to estimate interactively a background (*DMPLOT*, *WINPLOT*...). This is in fact generally the best way to proceed first, since trying to refine the background too soon may lead to instabilities when using pattern decomposition methods. The student may also try to estimate a space group automatically (*ABSEN*, *EQUIV*), but he will always have to verify manually or visually the result. Profile decomposition methods give the penultimate proof that the cell is correct. The session exercises will give the student the opportunity to verify this sentence.

Session 5 Become conscious of the main reasons for failure in SDPD is the subject of that session. One reason can be that preparing a pure sample is sometimes impossible. There is no equivalent to "one single crystal is enough". So that one should be able to cope with a multiphase powder pattern. The worst happens when several unknowns are mixed together. This is certainly the most complex situation and success was rarely obtained (one case is shown at the online conference "New Developments in Microstructure Analysis via Rietveld Refinement" [10]). Syntheses by varying the component concentrations often suggest that a sample is a mixture and allow to attribute some reflections to one phase and other reflections to another phase, due to intensity variations from one preparation to the other. But it is not always possible to completely eliminate a parasite phase, for instance when the process is hydrothermal synthesis. Fortunately, the parasite phase may be already known and will serve as an internal standard for indexing.

What should be the strategy, at the stage of structure factor amplitudes extraction, when a pattern of the pure unknown is unavailable ? The best is to take account of a known phase with structure constraint while the "|Fobs|" are extracted for the unknown. The student will have to deal with structure factor extraction in difficult conditions. Two patterns of mixtures in different proportions are provided for this exercise, allowing the student to apply his fresh whole knowledge.

Session 6 What can be done with those extracted structure factor amplitudes ? The session title is "Structure solution by Patterson, direct or molecule location methods, part 1 : Conventional methods." Many examples are available on the Web, including data and structure solutions. Software to download are those used for single crystal data : *SHELXS97*, *SIR97* (and you may want to control these programs by *WINGX*), *CAOS*, *CNS*, *CRUNCH*, *CRYSTALS*, *CSD*, *DIRDIF* (or *WINDIRDIF*), *GX*, *MITHRIL*, *JANA*, *MAXUS*, *MULTAN*, *SDP*, *SnB*, *TEXAN*, *UNICS*, *XLENS*, *GSAS*... Because the subject is powder data, some small programs may be useful in order to eliminate those "|Fobs|" that belong to overlapping reflections : *OVERLAP*. Programs include special powder diffraction data treatment (*EXPO*, *DOREES* in *POWSIM*, *FIPS*, *FOCUS*...), this will be the next session subject. Visualizing complete or partial models is requested at this session, by : *ORTEP* (or *WINORTEP*), *STRUVIR* (or *WINSTRUPLO*), *PLATON*, *CHIME*, *DRAWXTL*, *JAMM*, *JSV*, *ORTEX*, *POWDERCELL*, *RASMOL*, *XTAL3D*, *XTALDRAW*... A tremendous list of commercial software offer you drawing possibilities : *ATOMS*, *CRYSTAL MAKER*, *DIAMOND*, *SCHAKAL*, *CRYSTALLOGRAPHICA*, etc. You will find in *VRML* (Virtual Reality Modeling Language) a cheap way to access to three dimensional views of your models. A list of *VRML* tools for crystallography is available on the Web [11]. This session exercises will consist in solving the structures of the two compounds of which the structure factor amplitudes were extracted at sessions 4 and 5. The student should try to obtain a model, by using Patterson and/or direct methods, and Fourier recycling, which would further allow to start a refinement by the Rietveld method.

Session 7 Special methods for structure solution are examined in that session (molecule location excluded). We are close to the most recent developments in SDPD. Literature is scarce on those "special methods". Some may have been applied to a quite small number of real problems, and sometimes to none. There are very few programs available in the public domain, see the list at the SDPD-Database [8]. Moreover, the commercial programs can be expensive. For instance *POWDERSOLVE* from the MSI Company, a satellite program of the Cerius2 package, is only available to a consortium of the biggest pharmaceutical companies and to large scale facilities like ESRF. Fortunately, a few program under GPL (GNU Public License), with open sources, may be grabbed here and there. If you try to identify those special methods from the publication titles gathered inside the SDPD-Database, you will obtain the following list, as defined by the authors themselves, from 2000 back to 1991 : Monte Carlo from scratch; global optimisation method (*GOM*); general Monte Carlo approach; simulated-annealing method and a high degree of molecular flexibility; optimised data collection and analysis strategies; genetic algorithm; anomalous scattering difference; probability distributions for estimating the |F|s; computationally assisted; texture-based method; computer prediction; simultaneous translation and rotation of a structural fragment within the unit cell; combination of high-resolution X-ray powder diffraction and molecular modelling techniques; generalized rigid-body Monte Carlo method; solving crystal structures with the symmetry minimum function; static-structure energy minimization method; computer modelling approach; tangent formula derived from patterson-function arguments; real-space scavenger; bayesian approach; optimal symbolic addition program; entropy maximisation and likelihood ranking; Monte Carlo. Program names are available at the SDPD-Database [8] (*GAP*, *ROTSEARCH*, *OCTOPUS*, *ENDEAVOUR*, *DASH*...). Some packages exploit supplementary amount of information which become available during the phasing process : the preferred orientation, the pseudo-translational symmetry, the positivity of the

electron density, the positivity of the Patterson function, a well oriented and positioned fragment. Such an information allows theoretically to improve the pattern decomposition in *EXPO*. You may also download programs (*FOCUS*, *ZEFSA-II*...) that are devoted to solve framework structures (like zeolites). Software to download for that session are scarce : *EXPO* and *ESPOIR*, a brand new program [13] for structure determination by Monte Carlo from scratch (from a random starting model). The exercises consist in solving two structures by using these two programs.

Session 8 Chemists may know in advance which shape has a molecule, by NMR of a sample in a solution, or because of a well controlled synthesis process. Then, after indexing, the structure determination may consist in locating the molecule position inside the cell. Molecule location is the topic of that session 8. Possibly, there could be mediumly heavy atom (Cl or S) or small molecules (water, etc) which have to be simultaneously located, if the known fragment is not enough representative in percentage (80% may not lead to R factors below 30%) of the cell content. Many programs were built which perform rotations and translations of a model in the cell, until finding its correct position. When dealing with powder data, the problem is complicated due to reflections overlapping. Beside the brute force using a systematic grid-search approach, more elegant and efficient methods apply Monte Carlo/simulated annealing and also genetic algorithms in order to attain the best molecule position. The most elaborated programs may cope with several fragments simultaneously, and also explore torsion angles which may differ from the starting model. Review papers on that topic can be found at the SDPD-Database Web site [12]. That method for determining crystal structures is also used for large molecules, from single crystal diffraction data. The "Molecular Replacement" method is used for determining protein structures (*AMORE*, *MOLREP*, *MRX*, *MODELLER*... see the CCP4 Web pages [14]) or smaller molecules (*DIRDIF*, *PATSEE*). Some programs for structure prediction work by packing optimization : *PROMET*, *HARDPACK*, *PMC*, *UPACK*, etc. We are now not far from molecular modelling methods for optimization of models by semi-empirical or *ab initio* approaches. That session exercise will consist in solving an organic structure from a synchrotron powder diffraction pattern, knowing a large part of the molecule.

Session 9 The final SDPD step is reached : "Structure completion and refinement by the Rietveld method, part 1." When dealing with powder diffraction data, a structure should always be refined by using finally the Rietveld method. The original papers by Hugo Rietveld are online at his own Web site [15]. In the session 9 lecture, examples are treated showing how to complete and refine a structure by alternating the Rietveld method and Fourier difference syntheses. Clearly, the most popular Rietveld programs used by SDPD experts are freeware : *GSAS*, *FULLPROF* (avatar of *DBW*), *DBW* (distributed by R.A. Young), *WINMPROF*... New codes using the so called Fundamental Parameter (FP) approach are sometimes said to represent a revolution. Future will possibly confirm or not. Unfortunately, most programs using FP are commercial (*BGMN*, *TOPAS*...). Exercises will consist in completing some of the previous sessions structure determinations, but also in refining problematic data including undesired effects like preferred orientation and so on.

Session 10 "Structure completion and refinement by the Rietveld method, part 2" will more deeply introduce the student to difficulties. When and how to stop a Rietveld refinement are questions to which the student will be able to answer by himself only after years of experience. Should he refine the thermal parameters anisotropically, at least for some heavy atoms ? Should he try to find hydrogen atoms and should he include them in the refinements ? When are restraints and constraints applicable ? What is a good parameters/reflections ratio number with the Rietveld method ? Are the extracted size and microstrain parameters, possibly with anisotropic effects, dubious or not ? Etc. Acceptance or rejection of a manuscript will depend sometimes of such choices. The student should not go too far from reasonable limits, which are

certainly difficult to define exactly. Preferred orientation and anisotropic line broadening are maybe the two main problems (if the sample is pure, the resolution is the best as possible, and the statistics are fine) which could limit the quality of a refinement, and also could disallow to determine a structure from powder diffraction data. Finding hypothesis for hydrogen atom positions when Fourier difference maps do not show them is possible by using programs like *SHELXL*, *BABEL*, *XHYDEX*, *MOL2MOL*, *MOLDEN*, *WEBLAB*, *VMD*, *GROMACS*, *MOLDEN*... Optimization of a molecular model before going to refine with constraint/restraints or as a rigid body could be quite useful. This is a bit out of the scope of this SDPD course, but the student is invited to think to it. Programs well known by chemists can do that (*SPARTAN*, *GAUSS*...). For that session exercises, the student will have to complete the structure determination of an organic compound, proposing H atom positions. He will have to deal with a complex powder pattern including anisotropic line broadening. This is very probably the most difficult session.

Conclusion At the end of the course, the student will have completely determined the structure of a respectable number of unknowns by using the most powerful (easily available) software, and by applying very different approaches. For the final examination, the student will dispose of 2 weeks, starting at a date of his choice, for solving the structure of an ultimate unknown, by any method at his convenience, using powder diffraction data. Notation is made on 200 points for the final examination and 20 points for each of the 10 session exercises. Nothing is easy in this course, because SDPD is still not routine. Students should not expect to succeed without an heavy investment in the subject, abilities in computer and Internet access, as well as crystallography skills. Commercial codes, freeware, open source, those questions may have a strong impact on future education. The day when a teacher realizes that he will not be able to teach without the use of a commercial software is certainly a bad day. And this is what happen to us recently concerning the first week of the SDPD Internet Course : here was a bolt, and students not having access to a search-match program and ICDD database should directly go to the week 2 course material. Currently 9 students subscribed since September 1999, all over the world, all possessing a PhD or preparing it.

References

- [1] <http://sdpd.univ-lemans.fr/iniref/tutorial/indexa.html>
- [2] <http://sdpd.univ-lemans.fr/course/>
- [3] <http://www.cristal.org/sdpd/> - email to sdpd-subscribe@egroups.com
- [4] <http://crystalsun1.unige.ch/stxnews/riet/welcome.htm>
- [5] <http://sdpd.univ-lemans.fr/course/week-1/sdpd-1.html>
- [6] <http://www.cristal.org/SDPDRR/>
- [7] <http://www.ccp14.ac.uk/>
- [8] <http://sdpd.univ-lemans.fr/iniref/progmeth.html>
- [9] <http://sdpd.univ-lemans.fr/iniref/progmeth.html>
- [10] <http://sdpd.univ-lemans.fr/microstruct/denver/>
- [11] <http://sdpd.univ-lemans.fr/vrml/intvrml.html>
- [12] <http://sdpd.univ-lemans.fr/iniref/revpap.html>
- [13] <http://sdpd.univ-lemans.fr/sdpd/espoir/> - see also these EPDIC-7 proceedings.
- [14] <http://www.dl.ac.uk/CCP/CCP4/>
- [15] <http://home.wxs.nl/~rietv025/>