Direct space methods of structure determination from powder diffraction: principles, guidelines and perspectives

Radovan Černý*,1 and Vincent Favre-NicolinII

1 Laboratoire de Cristallographie, Université de Genève, 24, quai Ernest-Ansermet, 1211 Genève 4, Switzerland
II Université Joseph Fourier and CEA, DRFMC/SP2M/NRS, 17, rue des Martyrs, 38054 Grenoble Cedex 9, France

Received June 16, 2006; accepted August 24, 2006

Powder diffraction / Crystal structure solution / Direct space methods

Abstract. In the last twenty-five years structure determination from powder diffraction has evolved from a complex, time-consuming method to one easily usable. While the use of direct methods is restricted to samples well crystallized (sharp peaks, high resolution diffraction), structure solution using global optimization in direct space have become extremely popular, using either publicly or commercially available software. In this article we present a short history of direct space methods of structure determination from powder diffraction, and a list of available software. We then give an overview of frequent issues when solving structures in direct space, as well as how algorithms are currently evolving to solve more complex structures.

Introduction

Powder diffraction using X-rays and neutrons plays a major role in the search of new materials which are not available in the form of single crystals. Structure determination from powder diffraction (SDPD) is more difficult than structure determination on single crystals, because the available data are a projection of a three-dimensional diffraction pattern onto one dimension (radial distance from the reciprocal space origin), and consequently the diffraction peaks are overlapped. The extraction of structure factor amplitudes can be further complicated by a broadening (often anisotropic) due to the crystal lattice defects. Two alternative solutions of the peak overlap problem exist when trying to solve a crystal structure from powder data: either try to improve the decomposition of the observed powder pattern into individual peaks, or try to model the observed pattern as a whole. Consequently the methods of SDPD can be divided into two groups according to the working space (David et al., 2002 and references therein):

– Reciprocal space methods: They involve procedures developed for single crystal data like direct methods, Patterson synthesis or charge flipping (Baerlocher et al., 2006), and optimized for powder data. They need structure factor amplitudes obtained by powder pattern decomposition, and diffraction data up to 1 Å resolution.

– Direct space methods1: Different algorithms for a search in the direct space of structural parameters can be used, and an agreement factor between the observed and calculated powder diffraction data is evaluated, while the structural model is optimized to improve the agreement.

SDPD is also called ab-initio structure determination. One may wonder why such a name is used, the most common meaning for ab-initio in science meaning “from first principles”, i.e. with no model or data available, only physical laws. However for SDPD experimental data (the powder pattern) and basic model information is available! But the reason the ab-initio term has remained popular is because the both model information and experimental data are so incomplete, that most researchers in the powder diffraction community believe that their work starts (almost) “at the beginning”2.

In this article we will only deal with direct space methods – by this name we will not understand the interpretation of electron density or Patterson maps by search for molecular fragments, even if they work in direct space and use the global optimization algorithm like genetic algorithm in Chang, Lewis (1994). These methods still needs the structure amplitudes, i.e. decomposition of a powder pattern, which is avoided by direct space methods mentioned here.

---

1 From A. Le Bail talk on ESCA-9, Egypt 2004: Sometimes the “direct space methods” (not to be confused with the direct methods) are called “global optimization methods” or “model building methods”, and even sometimes “real space methods”. “Direct space” as opposed to “reciprocal space” has an adequate crystallographic structural sense, and should be preferred to “real space”, which, opposed to “imaginary” would call to mind both parts of the diffusion factors. “Global optimization” has a large sense and designates the task of finding the absolutely best set of parameters in order to optimize an objective function, a task not at all limited to crystallography.

2 See discussion in newsgroup: sci.techniques.xtallography in February 2001
Evolution of direct space methods

The first successful attempt to solve a crystal structure by an automatic (not manual!) localization of a building block (rigid molecule) in the direct space can be seen in the program RISCON (Bianchi et al., 1981) which was then modified for powder data as P-RISCON (Masciocchi et al., 1994). The optimization algorithm used was the constrained least-squares refinement, which is limited to structures not bigger than 10 independent atoms and resulted in only approximate atomic positions.

The authors of Pannetier et al. (1990) were one step from being the first using a true global optimization algorithm – simulated annealing (SA) for structure solution from powder data. However, they did not believe in the power of the method: “At present the method is not efficient enough for use in most practical problems of *ab-initio* structure determination.” The authors have used SA for structure prediction based on the optimization of the crystal potential energy. The first use of a global optimization algorithm (SA) in the structure solution from powder data therefore belongs to Newsam et al. (1992) even if the structure solved in the paper was known and small (benzene). Later on the direct space method of structure solution from powder data developed rapidly, using different algorithms like Monte Carlo (MC) search (Harris et al., 1994) and genetic algorithm (GA) (Shankland et al., 1997). An essential step forward was achieved by applying the description of structural blocks using internal coordinates like bond distances, angles and torsion angles in Andreev et al. (1997) allowing so a direct stereo-chemical interpretation and/or constraints on the optimized structural parameters. The modification of SA called parallel tempering algorithm (PT) was first used in the SDPD in Falcioni, Deem, (1999). Since that, the list of programs dealing with direct space methods of structure solution from powder (but also single crystal) X-ray and/or neutron diffraction data continues to grow. For the review see the Table 1 and http://www.cristal.org/ or http://www.ccp14.ac.uk/.

### Table 1. List of available computer programs which use direct space method for SDPD. Pawley refers to Pawley (1981), Le Bail refers to Le Bail et al. (1988).

<table>
<thead>
<tr>
<th>Program</th>
<th>Access</th>
<th>GO</th>
<th>CF</th>
<th>Reference</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>DASH</td>
<td>C</td>
<td>SA</td>
<td>P</td>
<td>David et al., 1998</td>
<td><a href="http://www.ccdc.cam.ac.uk">www.ccdc.cam.ac.uk</a></td>
</tr>
<tr>
<td>EAGER</td>
<td>A</td>
<td>GA</td>
<td>WP</td>
<td>Kariuki et al., 1997</td>
<td><a href="http://www.cardiff.ac.uk/chemistry/staff/harris.htm">www.cardiff.ac.uk/chemistry/staff/harris.htm</a></td>
</tr>
<tr>
<td>(former GAPSS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENDEAVOUR</td>
<td>C</td>
<td>SA</td>
<td>I + E</td>
<td>Putz et al., 1999</td>
<td><a href="http://www.crystalimpact.com">www.crystalimpact.com</a></td>
</tr>
<tr>
<td>ESPOIR</td>
<td>O</td>
<td>MC</td>
<td>L</td>
<td>Le Bail, 2001</td>
<td><a href="http://www.cristal.org">www.cristal.org</a></td>
</tr>
<tr>
<td>FOCUS</td>
<td>O</td>
<td></td>
<td>I + TS</td>
<td>Grosse-Kunstleve et al., 1997</td>
<td><a href="http://www.crystal.mat.ethz.ch">www.crystal.mat.ethz.ch</a></td>
</tr>
<tr>
<td>FOX</td>
<td>O</td>
<td>SA(P)</td>
<td>WP, I, AC</td>
<td>Favre-Nicolin and Černý, 2002</td>
<td>objcryst.sourceforge.net</td>
</tr>
<tr>
<td>OCTOPUS</td>
<td>A</td>
<td>MC</td>
<td>WP</td>
<td>Tremayne et al., 1997</td>
<td><a href="http://www.cardiff.ac.uk/chemistry/staff/harris.html">www.cardiff.ac.uk/chemistry/staff/harris.html</a></td>
</tr>
<tr>
<td>ORGANA</td>
<td>A</td>
<td>MC(E)</td>
<td>I + E</td>
<td>Brodski et al., 2005</td>
<td></td>
</tr>
<tr>
<td>POSSUM</td>
<td>A</td>
<td>DE</td>
<td>WP</td>
<td>Seaton and Tremayne, 2002</td>
<td><a href="http://www.chem.bham.ac.uk/labs/tremayne">www.chem.bham.ac.uk/labs/tremayne</a></td>
</tr>
<tr>
<td>POWDERSOLVE</td>
<td>C</td>
<td>MC</td>
<td>WP</td>
<td>Engel et al., 1999</td>
<td><a href="http://www.accelrys.com">www.accelrys.com</a></td>
</tr>
<tr>
<td>PSSP</td>
<td>O</td>
<td>SA</td>
<td>L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAFE</td>
<td>A</td>
<td>SA</td>
<td>WP + SE</td>
<td>Brenner et al., 2002</td>
<td><a href="http://www.crystal.mat.ethz.ch">www.crystal.mat.ethz.ch</a></td>
</tr>
<tr>
<td>SA</td>
<td>A</td>
<td>SA</td>
<td>WP</td>
<td>Andreev et al., 1997</td>
<td>ch-www.st-andrews.ac.uk/staff/pgb/group</td>
</tr>
<tr>
<td>TOPAS</td>
<td>C</td>
<td>SA</td>
<td>I, WP, E</td>
<td>Coelho, 2000</td>
<td>pws.prserv.net/Alan.Coelho</td>
</tr>
<tr>
<td>ZEFSAI1</td>
<td>O</td>
<td>MC(B)</td>
<td>I + AC</td>
<td>Falcioni and Deem, 1999</td>
<td><a href="http://www.mwdeem.rice.edu/zefsai1">www.mwdeem.rice.edu/zefsai1</a></td>
</tr>
</tbody>
</table>

Access: C = Commercial with academic prices, O = Open access, A = contact the authors
GO = Global Optimization: MC = Monte Carlo, MC(B) = biased Monte Carlo, MC(E) = Energy guided Monte Carlo, SA = MC + Simulated Annealing, PT = MC + Parallel Tempering, GA = Genetic Algorithm, DE = Differential Evolution
CF = CostFunction : P = Pawley, L = Le Bail, I = Integrated intensities, WP = Whole Pattern, E = potential energy, SE = structure envelopes, AC = Atomic Coordination, TS = Topology Search

### Principles of direct space methods

A flow chart representing a typical implementation, like in program Fox (Favre-Nicolin, Černý, 2002), of the global optimization approach to SDPD is given in Fig. 1.
At the end of the last century the direct space methods were developing intensively in the field of molecular (organic and metal-organic) crystals. Significantly less activity was found in the domain of non-molecular (inorganic) crystals. However, the same idea which was at the origin of the direct space methods success in the case of molecular crystals, was later applied also to the non-molecular crystals like extended solids or framework structures. The direct space methods are based (see Fig. 1) on location of building blocks (see chapter Model building) in the elementary cell by using random or systematic moves and/or modifications of the blocks, and comparison of calculated and observed diffraction patterns and/or other cost functions (CF) like crystal energy, atomic coordination etc. Based on a “fitness” of the current structural model the decisions are taken how to improve the model. Generally said, it is a global optimization problem of a great complexity, where the algorithm must explore a hypersurface (see Fig. 4) which describes the “cost” of the model as a function of all structural parameters (see chapter 15.6 in David et al., 2002), and find its global minimum. Two algorithms of global optimization have found larger application:

Simulated Annealing (SA) or Parallel Tempering (PT)

SA and PT algorithms are both based on MC sampling, earlier known as “statistical sampling” (for review, see Newman, Barkema, 1999). The first, and now widely used, algorithm of the MC sampling is based on Boltzmann distribution, and is known as Metropolis algorithm (Metropolis et al., 1953). The MC sampling as applied in SDPD is also called Reverse Monte Carlo (McGreavy, Pusztai, 1988), because the system is modified by random changes under the constraint of observed data, like a diffraction pattern. A general flow chart of the Metropolis algorithm applied to SDPD is given in Fig. 2.

The PT algorithm was originally introduced in the context of spin glass simulations (Hukushima, Nemoto, 1996), similar ideas were also proposed in Geyer (1992). Its principal advantage within the SDPD compared to SA algorithm is its generality for any type of problem; no parameters like annealing rate, starting temperature, are required. The algorithm is also generally able to escape from the local minima in the parameter space (Favre-Nicolin, Černy, 2002).

Evolutionary theory – Genetic Algorithm (GA)

Computing using GA, a fast developing tool of science and technology, is helpful in solving many-parameter optimization tasks. GA form a subset of broader classes of global-optimization strategies called population-based methods, and evolutionary algorithms. The concept of GA follows the old idea of minimizing human efforts in solving difficult scientific and technical problems by learning from the nature (see Holland, 1975, Goldberg, 1989). The genetic computation proceeds in the space of (usually binary coded) variables. It mimics the evolution of living organisms represented by the points in this space (trial solutions). In the beginning, a population of individuals (called also chromosomes, agents...), which may represent trial solutions of the optimisation task, is generated. Next generations are successively created using simplified principles of plant or animal (Darwinian) evolution. The calculation is terminated by application of a suitable stop condition. The basic genetic operators used in formation of each new population include selection, crossover and mutation. GA was first used for SDPD in Shankland et al. (1997) and Kariuki et al. (1997). A general flow chart of the genetic algorithm applied to SDPD is given in Fig. 3.

Important points for successful structure solution

In spite of the fact that excellent software packages exist (see Table 1), it is frequent that structures fail to be solved, even when outstanding data is available. In this part we will detail the steps to take to maximize the chance of solving the structure.

---

**Fig. 2.** A typical flow chart of the Metropolis algorithm as applied to SDPD, case of the program Fox (Favre-Nicolin, Černy, 2002).

**Fig. 3.** Flow chart of the genetic algorithm as applied to SDPD, general case.
Gathering maximum prior information

First of all, it is vital to gather as much prior information on the sample before even collecting experimental data. The most basic information is the chemical formula of the compound, and if possible, for molecular compounds, a plane or spatial formula. For non-molecular crystals, the knowledge of the structure of similar compounds (what polyhedra it forms, inter-atomic distances).

One should always double-check the chemical information, i.e. often the announced chemical formula can turn out to correspond to the desired compound expected by the chemist (nominal composition), and not the actual one (measured or refined composition). A quick chemical analysis using suitable spectroscopic technique (EDAX, atomic spectroscopy, mass spectroscopy etc.) can be extremely useful to avoid looking for the wrong compound. Finally, the possibility of solvent molecules in the crystal structure should always be considered.

Another information which can be gathered using microscopic techniques is the shape of the crystallites: if they have a needle or planar shape, this is the sign of probable preferred orientation, which must be avoided at all cost in the next step unless one plan to use it for resolving the separation of overlapping reflections in the powder pattern when using the reciprocal space methods (Baerlocher et al., 2004).

NMR data can also be used to determine the plane formula for molecular compounds or the type of polyhedra present in the material for non-molecular samples (Edgar et al., 2002).

Gathering maximum experimental data

The first data gathered is always from a laboratory diffractometer – modern instruments allow to measure high-quality diffraction data with FWHM down to 0.05 deg, which is sufficient to solve many structures. In order to minimize the risk of having preferred orientation effects, it is recommended to use samples prepared in a capillary rather than in a flat plate unless the highly absorbing samples require the reflection geometry. For more guidelines on optimal data collection, see McCusker et al. (1999).

When working on a complex sample (more than 20 independent atoms, or unit cell vector larger than 15 Å), it can be helpful to collect higher data quality using synchrotron radiation, which can yield a higher flux (lower noise in the data), sharper peaks with FWHM < 0.005 deg (less overlapping reflections) and lower background when using crystal analyzer (detection of weak reflections allowing to find correct cell). However this can be difficult for samples which are radiation-sensitive and can be quickly damaged under the high synchrotron flux – fast data collection using multi-detectors (Hodeau et al., 1998) or strip detectors (Schmitt et al., 2003) have been recently implemented to allow data collection before the structure collapse of the sample.

Finally, neutron diffraction can be considered – as the neutron scattering length is fundamentally different from the X-ray one, it allows to outline lighter atoms (most particularly deuterium), and to distinguish neighbouring elements whose ions can share the same number of electrons (such as Zn\(^{2+}\) and Cu\(^{+}\)). In the case when sufficient amount of sample cannot be gathered for neutron diffraction, an alternative method to distinguish neighbouring elements is resonant diffraction close the absorption edge of both of the elements (Zhang et al., 2003). For diffraction data, more data gathered is always better, as it is possible to use joint optimization on several data (e.g. neutron + X-ray) sets to maximize the chance of finding quickly the true solution.

Space group and unit cell determination

Once diffraction data has been obtained, the first step towards the structure solution is the determination of the unit cell (indexing) and space group. This can be a complex and sometimes is the most difficult step for ab-initio structure determination, and is the subject of abundant literature (Bergmann et al., 2004). Practically, tools such as Crysfire (Shirley, 2002) will yield a list of possible unit cells and space groups with a ranking, generally giving few possibilities which can then be used for the algorithm performing the structure search in direct space.

The determination of the space group is theoretically not necessary, as the structure can be searched in space group P1, however, as the number of trials required to find the true solution increases exponentially with the number of independent atoms, a factor 2, 4, 8 up to 192 reduction in the number of atoms cannot be passed on. In the extreme case when no unit cell or space group candidate can be obtained, it is still possible to do a direct space determination using the Pair Distribution Function (PDF) modeling (Juhás et al., 2006). This analysis does not require long-range periodicity and symmetry (but does not exclude it), and is sensitive to the short-range (local) order (Egami and Billinge, 2003).

Model building

This is a most crucial step: any direct space search requires the definition of the degree of freedom (DoF) which describes the structure (see Fig. 4). The most simple approach is to determine the approximate number N of independent atoms, and then use the 3*N crystallographic coordinates as DoF; of course the simple approach is the large number of DoF.

Any crystal structure can be described as a combination of scattering objects — building blocks, which can be independent atoms, molecules, polyhedra or molecular fragments like amino-acids. The volume of the unit cell can be used to determine the number of independent building blocks in the asymmetric unit from the known crystal density or average volume occupied by a single atom (around 15–20 Å). The advantage of this description is that by a suitable choice of geometrical descriptors that characterize the crystal structure, the number of DoF can be decreased.

Molecular crystals

The building blocks based description is particularly efficient for molecular materials, where molecules are natural blocks, planar and sometimes spatial formula of the mole-
The final choice is not trivial, particularly due to the presence of special positions. The use of a dynamical occupancy correction (DOC) (Favre-Nicolin, Černý, 2002) simplifies the problem as it is no longer necessary to manually adjust the occupancy of atoms falling in a special position, and also makes two identical elements fully overlapping half-occupied to be “seen” by the diffraction as a single atom. DOC has revealed to be very powerful in the case when the exact composition of the studied compound is a priori not known exactly, like hydrogen in metal hydrides obtained by the hydrogen absorption in a metallic matrix of an alloy or intermetallic compound (for examples see Černý, Favre-Nicolin, 2005).

Theoretically, this means that it is possible to add more atoms than initially deemed necessary, expecting the DOC to artificially “merge” the excess atoms. In practice, adding too many atoms will slow down the optimization, as more atoms have to fall in the correct position at the same time to find the correct structure. This also implies that if some atoms are shared between polyhedra, a choice should be made on where the bridging atoms should be used.

A nice example were careful model building can help can be found in Edgar et al. (2002), the structure solution for \( \text{Al}_2(\text{CH}_3\text{PO}_3)_3 \) was done using two different models: using (i) \( 2\text{Al} + 3\text{CH}_3\text{PO}_3 \), and (ii) \( \text{AlO}_4 + \text{AlO}_5 + \text{CH}_3\text{P} \). Both models avoid the inclusion of oxygen atoms in both the \( \text{AlO}_5 \) polyhedra and the phosphonate molecule, to keep the number of independent oxygen atoms equal to 9. However model (i) allowed finding the correct solution in 750 \( \times 10^3 \) trials while (ii) required 6.5 \( \times 10^6 \) trials. This can be explained by the presence in model (i) of equivalent building blocks which lowered the conformation space to search. Moreover, in model (ii) not only are the \( \text{AlO}_4 \) and \( \text{AlO}_5 \) blocks independent, but they are very similar which can easily create false minima which can slow down the optimization.

Structure validation

There are a number of statistics which can be used to assert how good is the agreement between observed and calculated diffraction data, the most commonly used being the \( \chi^2 \) and \( R_{wp} \) (Toby, 2006). However, these must be used with a care. A “good” \( R_{wp} \) value will depend a lot on the quality and type of powder pattern recorded—a 12% may be very good for a synchrotron powder pattern with a very low noise, but would be very bad with neutron or laboratory data with a high background. It is also possible to compare the finally obtained \( R_{wp} \) to the value obtained using profile-fitting (David, 2004).

In practice, the best tool to evaluate how good a fit is still the eye of an experienced crystallographer (McCusker et al., 1999), but of course other criteria have to be taken into account, and most particularly chemical ones:

- Are there atoms “colliding” each other? If the model repeatedly finds a global solution with atoms overlapping, an anti-bump penalty (see for example Favre-Nicolin, Černý, 2002) can be used to avoid this solution. Note that this should only happen if the quality of the powder pattern is insufficient, and therefore includes insufficient information to “lock” the algorithm to the correct solution.
- Do the atomic environments make sense? A penalty based on bond valence sum calculation (Pannetier et al., 1990) can point out unlikely configuration. Again, if valences are known a priori, this can be used as a criterion for the optimization of the structure.

Macromolecular crystallography often uses the so-called \( R_{max} \) factor for single crystal refinements (Brünger, 1992). A percentage (typically 5%) of reflections are set
aside and not used during the refinement. To test the validity of the model, the $R$-factor can then be calculated using only the excluded reflections. This test is particularly useful to avoid false minima. Of course for powder diffraction setting aside a sufficiently large number of reflections to be used for statistics is a luxury given the small number of reflections generally used. However, the direct space method algorithms only use relatively low-resolution data (up to 2 or 1.5 Å). So after a solution has been proposed, it is possible to extend the range over which the observed and calculated patterns are computed. If the solution is correct, the powder patterns should still fit immediately after the limit used for the structure solution optimization.

The best validation of the structural model resulting from the diffraction data would be an agreement with a structural model predicted either by a quantum-chemical stability calculation (Dovesi et al., 1996), combined with local or global energy minimization (Beyer et al., 2001; Schön, Jansen, 2001) or by an application of crystallographic rules derived from crystal chemistry of known structures (Le Bail, 2005). However, this belongs rather among the challenges and perspectives.

Current challenges and prospects for SDPD

Preferred orientation identification

Preferred orientation is often the greatest problem when trying to solve a structure. Although it is possible to evaluate *ab-initio* the preferred orientation (Altomare et al., 1994), this is still rare. Theoretically, one can treat the texture parameters like any other parameter in the optimization, using the March-Dollase model (Dollase, 1986), this only requires 3 parameters: the \([hkl]\) coordinates of the preferred orientation vector (only 2 independent parameters), and the March coefficient describing the type (plate or needle) and intensity of the preferred orientation. But in practice including these parameters will require 10 to 100 time longer runs, which can be unacceptable.

An intermediate solution would be if the preferred orientation direction was known in advance, which is often the case when the unit cell is strongly anisotropic and if microscopy has been used to study the shape of the crystallites. Then the only remaining parameter would be the March coefficient, which could even be refined on each trial by the algorithm.

Using flexible structure models (which still satisfy the user)

Designing a model and choosing the DoF to be optimized is only the “visible” part of the structure solution building. While reducing the number of DoF using strict restraints is a good idea, putting too many restraints can prevent the algorithm from converging. *e.g.* imagine an organic molecule with a long (>8 atoms) ring: if strict restraints are set on all the bond angles, it then becomes impossible to alter the conformation of the ring. This can even be worse if it is necessary to flip the position of one side group.

The consequence of this is that while the structure solution program must allow for a large number of restraints to restrict the parameter space being explored (and the user must be encouraged to use them), the algorithm must be allowed to violate these restrictions when it can allow a better model to be found. Needless to say, this is a source of confusion between users and algorithm designers (“Why does the program not respect my restraints?”).

In Fox (Favre-Nicolin, Černý, 2002), restraints (bond lengths, bond angles and dihedral angles) can be used and will remain “flexible”, *i.e.* the algorithm is allowed to have a number of restraints going far from the assigned value. But it is still possible to define sub-units of a molecule which remain strictly rigid.

Meta-model building

Choosing a correct model for inorganic samples is difficult because of the many ways a structure can be described, especially when the only knowledge about it is an approximate chemical formula. A structure with 6 atoms in the unit cell can require a single atom in a general position, or 6 in special positions!

This is particularly hard to resolve, as the nature of diffraction is such that the absence of a minority of atoms in the model (10–20%) can make the entire model look wrong for the diffraction — thus it is quite hard to determine “partial models” even if they are properly described based on maximum likelihood (Markvardsen et al., 2002).

This could be improved using meta-model building, *i.e.* instead of saying that the model is “$P_2_1$ space group, with 2 PO$_4$ tetrahedra and 6 WO$_6$ octahedra”, the user could indicate that the optimization should use “either $P_2_1$ or $P_2_1/a$ space group, with between 1 and 5 PO$_4$ tetrahedra and between 3 and 8 WO$_6$ octahedra”. It would then be up to the software to, either change the space group and number of building blocks during the optimization, or to make a series of runs with different models.

Distributed computing

While computer power has been steadily increasing for a long time, it is still limited and structure optimization can require months (Peschar et al., 2004) to solve the most complex structures. But more computing power can be achieved using not a single computer, but a “grid” of computers. In the last 10 years a large number of large computing projects have emerged which require extraordinary amount of computation, such as the seti@home and climateteprediction.net programs. These are now grouped in the BOINC project (http://boinc.berkeley.edu/). These projects all make use of “idle cycles” on computers all around the world — projects can then use at a very low cost 100’s of thousands of computers.

Of course such a computing power is not necessary for day-to-day SDPD. But small grids (5–20 computers) could be extremely useful for complex optimisations, such as when preferred orientation is present or if several models are available. This has already been used to enhance *ab-initio* SDPD algorithms (Markvardsen et al., 2005). Future software should include the possibility of distributing
optimization jobs – this would enable users the use of less restricted models without the fear of it running too slowly.

Towards better optimization algorithms

Many algorithms are available within the “global optimization community” (see for example Floudas, Pardalos, 2000). However, only the incomplete algorithms, i.e. without any guarantee to reach the global minimum (Neumaier, 2004) like SA, PT and GA have been used within SDPD. The new developments include hybrid algorithms like hybrid MC (Johnston et al., 2002).

Including genetic algorithm principles

Right now no software package using GA for SDPD is publicly available (either freely or commercially). However, several studies have shown its powerful possibilities (Harris et al., 2004). The implicit parallelism of the GA is its greatest advantage and attracted many developers to import this property in the MC sampling based algorithms like SA and PT to improve their performance and make them suitable for parallel computing. Several hybrid algorithms have appeared within the global optimization community, for review see Mahfoud, Goldberg (1995).

Maximum Likelihood (ML)

ML (for an introduction see Sivia, 1996) has been used with a lot of success for macromolecular crystallography (Read, 1990; Pannu, Read, 1996; Murshudov, Vagin, Dodson, 1997; Read, 2001). During a normal (Rietveld) evaluation of a structural model, all atoms must have a clearly defined position and occupancy to calculate a powder pattern and compare it to the observed one. Using ML allows evaluation of approximate structural models, where not all atoms have well-defined positions.

This is achieved by computing not only the calculated diffracted intensities but also the uncertainty (as the variance $\sigma^2$ associated with these intensities, which reflects the uncertainty in the model. This variance is then combined with the one on the observed powder pattern, and is used in the calculation of the $\chi^2$.

This approach is a boon for SDPD since by definition, when searching for a structure solution the atomic positions can only be known in an approximate manner. This was used first by Markvardsen et al. (2002) to allow solving the crystal structure of an organic compound while keeping the position of solvent molecules completely unknown. More recently it was shown (Favre-Nicolin, Černý, 2004) that it could be generally used to improve the convergence of direct-space algorithms by introducing a systematic positional error for the atoms.

Fox: Free Objects for Xtallography

Since its release in 2001 the program Fox (Favre-Nicolin, Černý, 2002) has been quite often used for solving crystal structures from powder diffraction data (for a list of solved structures see http://objcryst.sourceforge.net/Fox/, for a review of non-molecular structures see Černý, Favre-Nicolin, 2005). The complexity of structures solved by Fox ranges from 2 to 34 independent atoms found ab-initio for inorganic compounds. Decreasing the DoF by modelling the structure with larger building blocks was one of the reasons of using Fox for SDPD of non-molecular compounds. In many cases the tetrahedral and octahedral units were successfully used. For organic compounds, the largest structure solved included 59 non-H (and 102 H) atoms. A list of structures solved with Fox can be found on the Fox wiki http://objcryst.sourceforge.net/Fox/FoxBiblioStructures.

Fox is open-source software, released under the GNU General Public License. It can be downloaded from http://objcryst.sf.net/Fox Precompiled versions are available for windows and MacOS X. Future developments of Fox do not depend only on its authors, but as Fox is available as open source program, also on any user who decides for modifications.

Conclusions

Structure determination from powder diffraction data has undergone an intensive development in last 25 years. First applying the reciprocal space methods, which were optimized to work with lower quality data obtained from powder diffraction patterns, and with no important difference when applied to molecular or non-molecular crystals. The direct space method has rapidly evolved within last 15 years, and continues still to be developed, as user-friendly tool for SDPD. The main principles are same for molecular and non-molecular crystals, however, some specials tools were necessary to develop for treatment of special crystallographic positions, sharing of atoms between different building blocks like coordination polyhedra, and for correct optimization of disordered atomic positions.

Current (known) limits of direct space methods are around 30–100 independent atoms, depending on the symmetry and the known constraints in the structure. The success depends on the quality of the diffraction data and even more on the amount of the additional chemical information (knowledge about structural building blocks) injected into the structure solution process. The same is true for the reciprocal space methods. However, the use of the additional information, like atomic coordination, interatomic distances, angles, is easier and natural when working in direct space, the space where this information comes from.

Among the current challenges and prospects of SDPD belongs active modelling of preferred orientation, active evolution of the structural model during the optimization, improvement of the optimization algorithm, and speeding-up the calculation. All these developments may proceed towards an automatic SDPD connected with the structure prediction. Actual state of knowledge, however, still requires active interaction of an experienced crystallographer.

Acknowledgments. The authors want to thank to all users of Fox, and especially to them who have kindly provided the details of their work when solving the crystal structures. The discussion with Yuri Andreev from the University of St. Andrews on the principles of direct space methods is highly appreciated. The review of Genetic Algorithm principles relies much on the discussion with Wojciech Paszkowicz from the Polish Academy of Sciences in Warsaw, which is highly appreciated.
References


