Direct Space Approaches

Rosanna Rizzi
Institute of Crystallography, Bari, Italy
rosanna.rizzi@ic.cnr.it

Introduction

The Direct Space (DS) methods have been successfully applied for solving polycrystalline structures, provided the expected molecular geometry is known. Avoiding the extraction of the integrated intensities from the experimental powder pattern, a so critical step for the success of the Reciprocal Space (RS) methods (like Direct Methods), they can succeed in case of low resolution and/or complex structure and/or bad quality data, for which, at the contrary, RS methods often fail. The DS approaches work in direct space where trial structure models, built by using prior information about the expected molecular geometry, are adjusted in terms of position, orientation and internal conformation in order to find the best set of structural parameters which minimize a cost function (CF), usually the agreement between the measured and calculated diffraction profile (R_{wp}). The structure solution is then equivalent to searching for, and locating the position of, the minimum value of an *N*-dimensional hypersurface (where *N* is the number of structural parameters to be varied). Grid Search [1], Monte Carlo [2-4], Simulated Annealing [5] and Genetic algorithm [6-8] are the search methods most commonly used for this purpose and they will be described in the following session together with Parallel Tempering [9] and the Big Bang-Big Crunch approach combined with Simulated Annealing [10]. A brief description of the hybrid methods will be also given.

Direct Space Approaches

Grid Search

The most intuitive approach to handle the crystal structure solution by DS techniques is to define a grid over the parameter space of interest and to perform an exhaustive grid search. This method, simple and easy to be implemented in a computer code, ensures that the global minimum will be found providing that a suitably fine grid is used. This requires long computing time to complete the procedure. For this reason, Grid Search can be successfully applied to some problems in powder diffraction where the parameter space is relatively small, *i.e.*, when the number of degrees of freedom (DOFs) necessary to describe the model, is limited. The Grid Search procedures have been mainly applied to rigid structural fragment.

Monte Carlo Methods

To obtain a good quality crystal structure solution in a reasonable amount of time, Grid Search methods should be replaced by the stochastic ones, based on a random sampling of the parameter space: the Monte Carlo (MC) technique. MC methods require the construction of a Markov chain of structures, *i.e.*, a sequence of structures each exclusively depending on the previous one. Each structure of the chain is described by a set of external and internal DOFs necessary to fix the position, the orientation and the internal conformation. The MC steps can be summarized as follows:

1) A new configuration is derived from the previous one by random variations of its DOFs.

2) A powder diffraction pattern is calculated from the trial structure and compared with the experimental profile. The quality of the agreement is generally evaluated by the CF representing the profile weighted index $R_{wp} = \sqrt{\frac{\sum_i w_i \left(y_i^o - y_i^c\right)^2}{\sum_i w_i \left(y_i^o\right)^2}}$ (1)

where y_i^o and y_i^c are the observed and the calculated counts corresponding to the *i*-th point in the powder diffraction profile, respectively, and $w_i = 1/y_i^o$.

3) An acceptance criterion decides whether the trial structure is accepted or not. Generally, the Importance Sampling [11] algorithm is used for this. The trial configuration is accepted if $CF < CF_{old}$ otherwise it is accepted with probability:

$$\exp\left[-\left(CF - CF_{old} / T\right)\right] \qquad (2)$$

3) If the trial configuration is accepted, it becomes the starting point for a subsequent generation otherwise, a new generation is attempted from the old configuration until the acceptance criterion is fulfilled.

Several MC moves are performed until a sufficiently long chain of structures is obtained; this can be viewed as a random walk through the parameter space.

Simulated Annealing procedure

The fundamental difference between standard MC and Simulated Annealing (SA) techniques concerns the way in which the scaling factor *T* of Eq. (2) is used to control the sampling algorithm. In the MC method, *T* is fixed or varied empirically, whereas in the SA *T* is slowly reduced during the procedure, according to an annealing schedule as follows:

1) The system starts at a given *T* value, *T* is then increased until the number of accepted configurations is high enough. This ensure the system is melt, *i.e.*, the DOFs can vary freely in the parameter space;
2) *T* is decreased, keeping the same value until a fixed number of moves has been performed;
3) The reached global minimum position is refined by a rapid decrease of *T* and the iteration is stopped when the acceptance ratio reaches some allowed low value, indicating that no further improvements are observed.

Several variants of the basic SA algorithm have been proposed, which differ in the choice of the CF, in the design of the annealing schedule, etc.

Parallel Tempering

An improved version of the SA search algorithm, called Parallel Tempering, has been developed by Favre-Nicolin [12]. Since a single chain of configurations could be trapped in local minima if the temperature decreases prematurely, a small number of parallel optimizations are performed, each at a different temperature. Exchanges of configurations among parallel optimization processes are allowed, using the same acceptance criterion reported in the Eq. (2), to ensure an optimal exploration of the parameter space. This approach is particularly suited for structure models defined by a high number of DOFs.

Genetic Algorithm Techniques

The Genetic Algorithm (GA) is an alternative global optimization technique to MC and SA approaches. It is based on the principles of the Darwinian theory of evolution and allows a

population composed of many individuals to evolve under specified selection rules to a state of best fitness. The operations involved in a single evolution step are:

1) *Natural selection*. It selects the individuals destined to survive or to take part in mating. It drives the evolution towards the best fitted population;
2) *Mating*. It procreates the new offspring by mixing the genetic information of two selected parents. This step is responsible for the preservation and improvement of selected characters in the population;
3) *Mutation*. It generates new individuals (mutants) by randomly changing part of the genetic information of individuals randomly chosen from the population. Mutation is responsible for genetic diversity and prevents the stagnation of the population.

A relevant feature of GA methods is its implicit parallelism, being able to treat at the same time different members of a population, so investigating simultaneously different regions of the parameter space. This makes the GA particularly advantageous for the solution of problems with many DOFs.

***Hybrid* Big Bang-Big Crunch**

The *Hybrid* Big Bang-Big Crunch method (*H*BB-BC) results from a proper combination of Big Bang-Big Crunch (BB-BC) algorithm [13] with *SA*. BB-BC relies on one of the evolutionary theory of the universe consisting of two successive phases: 1) the Big Bang, corresponding to an energy dissipation procedure for creating a completely random initial population; 2) the Big Crunch, corresponding to a contraction procedure for converging to a global optimum point.

The main steps of *H*BB-BC are the following:

a) An initial random population is created (first Big Bang) and for all its members the reliability parameter reported in the Eq. (1) is calculated;
b) The Big Crunch phase follows by choosing the candidate of the population with the lowest R_{wp} value as centre of mass (x_c);
c) The Big Bang explosion is carried out again generating, around x_c, new solution candidates;
d) The R_{wp} value is calculated for all the candidate models and, at most three models with the worst R_{wp} values, are chosen and submitted to performing a fast annealing schedule;
e) The procedure returns to the step b) and is cyclically repeated until a stopping criterion is reached.

The application of this procedure to a large variety of powder compounds has clearly shown its ability, respect to the traditional SA algorithm, to gain computing time without losing the number of determined correct solutions.

The Hybrid Approaches

The more recent developments of global optimization approaches, have led to hybrid methods [14-16]. Developed and implemented in some computing programs, they consist of the combination of both RS and DS methods, trying to exploit the best features of the two approaches. Examples of largely used hybrid algorithms are:

1) the Charge Flipping Method (CFM) [17]. CFM is a dual-space iterative structure solution method switching back and forth between real and reciprocal spaces. It starts from an electron density map calculated by the observed structure factor moduli and random phases. Then the map is modified in the direct space by reversing (flipping) the sign of all the densities below a given small positive

threshold. The procedure is cyclically applied and stops when the convergence is reached;

2) DS approaches combined with *ab initio* methods. Some located positions of the model, as obtained after the application of the traditional Direct Methods procedure, are used as pivots on which the expected structural model is settled. This strategy reduces the number of degrees of freedom of the problem in the optimization process (no translation is involved) so making the space exploration faster [18-19].

Conclusions

In the last 25 years, the structure solution from powder data has been characterized by the intensive development of new algorithms, which have made easier the crystal structure solution. Powder compound with 30-40 atoms in the asymmetric unit can be now solved. Direct Methods advances have been accompanied by the strong evolution of Direct Space and Hybrid approaches.

References

[1] Chernyshev, V.V. & Schenk, H. (1998). *Z. Kristallogr*. **213**, 1-3.

[2] Andreev, Y.G., Lightfoot, P., Bruce, P.G. (1997). *J. Appl. Cryst.* **30**, 294**–**305.

[3] Harris, K.D.M, Tremayne, M., Lightfoot, P., Bruce, P.G. (1994). *J. Am. Chem. Soc*. **116,** 3543**–**3547.

[4] Tremayne, M., Kariuki, B.M., Harris, K.D.M., Shankland, K., Knigh, K.S. (1997). *J. Appl. Cryst*. **30**, 968**–**974.

[5] Kirkpatrick, S. (1983). *J. Stat. Phys.* **34**, 975**–**986.

[6] Goldberg, D.E. (1989). In *Search, Optimization, and Machine Learning*. New York, USA: Addison-Wesley.

[7] Kariuki, B.M., Serrano-González, H., Johnston, R.L. & Harris, K.D.M. (1997). *Chem. Phys. Lett.* **280**, 189**–**195.

[8] Shankland, K., David, W.I.F., Csoka, T. (1997). *Z. Kristallogr*. **212**, 550**–**552.

[9] Favre-Nicolin, V., Černý, R. (2004). *Z. Kristallogr*. **219**, 847-856.

[10] Altomare, A., Corriero, N., Cuocci, C., Moliterni, A., Rizzi, R. (2013). *J. Appl. Cryst.*, **46**, 779-787.

[11] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E. (1978). J. Chem. Phys., **2**1, 1087.

[12] Favre-Nicolin, V., Černý, R. (2002). *J. Appl. Cryst.*, **356**, 734.

[13] Erol, O.k. & Eksin, I. (2006). *Adv. Eng. Software* **37**, 106-111.

[14] Johnston, J.C., David, W.I.F., Markvardsen, A.J., Shankland, K. (2002). Acta Cryst. A**58**, 441-447.

[15] Brenner, S., McCusker, L.B., Baerlocher, C. (1997). J. Appl. Cryst. **30**, 1167-1172.

[16] Altomare, A., Caliandro, R., Giacovazzo, C. et al. (2008). J. Appl. Cryst. **41**, 56**–**61.

[17] Oszlányi, G. & Sütó, A. (2004). *Acta Cryst*. A**60**, 134-141.

[18] Altomare, A., Caliandro, R., Giacovazzo, C., Moliterni, A.G.G., Rizzi, R. (2003). *J. Appl. Cryst.* **36**, 230-238.

[19] Altomare, A., Caliandro, R., Cuocci, C., Giacovazzo, C, Moliterni, A.G., Rizzi, R, Platteau, C. (2008). *J. Appl. Cryst.* **41**, 56-61.