Monte Carlo indexing with McMaille

Armel Le Bail

Université du Maine, Laboratoire des Fluorures, CNRS UMR 6010, Avenue O. Messiaen, 72085 Le Mans Cedex 9, France - E-mail : alb@cristal.org - Web : http://cristal.org/

The McMaille (pronounce MacMy) computer program code [1] was written up by the end of 2002, in order to explore the Monte Carlo possibilities for indexing powder patterns, as suggested in a recent paper (Kariuki *et al.*, 1999) [2] proposing a new approach based on whole-profile fitting and global optimization using a genetic algorithm. Computer power and speed have progressed enormously, so that some time-consuming approaches have become more practicable. Rather than to retain whole-raw-profile fitting, a first McMaille version tried to fit a pseudo powder pattern built up by using Gaussian peak shapes having full widths at half maximum (FWHM) following the (*U*, *V*, *W*) Caglioti law characterizing standard patterns from the used diffractometer, the peaks having positions and intensities obtained from the application of a peak hunting software (PowderX or WinPlotr for instance). However, it was found that a Monte Carlo process, that will randomly propose cell parameters, would have better chances of success if the FWHM were enlarged rather than narrowed. This does not mean at all that data would not have to be accurate. On the contrary. This just means that using enlarged peaks, centered around a very accurate position, would give more chance to the process to detect quickly a minima in the figure of merit (FoM) surface, starting from cell parameters decisively more different of the final ones than if the FWHM were too narrow. As FoM, the conventional Rietveld R_P value was retained. A problem was that the yet simple Gaussian peak shapes, combined with 3-4 iterations with the Le Bail method for fitting the pattern, needed too much computer time (ref. [2] gives no idea of the time needed for indexing a small orthorhombic cell). By using a computer running McMaille at 2.4 GHz, on a fragment of pseudo powder pattern built up from 20 peak positions and intensities, it was possible to test 10^3 cells per second in cubic symmetry, much less in lower symmetries. Then, even more simple columnar peak shapes were tested, no Le Bail fit , but the *R* factor was estimated from the percentage of inclusion of the calculated column inside of the "observed" one (of course, the calculated column intensity was set equal to the "observed" one). The calculations were 20 time faster, leading to possibilities for indexing in any crystal system in more reasonable times (in a matter of seconds for high symmetry and minutes for low symmetries including triclinic in less than 1/4 hour). However, such times are relevant to the examination of a restricted domain of volume (ΔV = 500 Å^3) and of cell parameters (< 20 Å). Examining all symmetries in a quite large domain requires hours, if not a night of calculations, testing up to 10^9 cell parameters combinations. Four tricks have a part in the success of the Monte Carlo process, changing randomly one parameter at a time, that parameter being itself selected randomly, depending on the symmetry (from 1 to 6 parameters, zeropoint fixed) :

a - Cells are retained for further examination if *R* is smaller than a user defined value (~50%).

b - Cells are also retained for further examination if all the observed peaks (minus a number of tolerated impurity peaks defined by the user) are "explained", whatever the *R* value.

c - Further examination means that if a) or b) conditions are fulfilled, then the cell parameters are adjusted by a Monte Carlo process, testing randomly 200 to 5000 small parameter changes (cubic to triclinic case, respectively). That way, *R* can decrease from 50% (case a) or larger (case b) to the minima (usually less than 10%), which a least-square refinement process would not have allowed.

d - Memory is kept of new parameters if they improve *R* in 85% of the cases (in order to escape from false minima). Tests of efficiency of the process were made for various percentages.

The relative insensitivity to impurities is a strong point of McMaille. Tests for indexing simultaneously two phases in a mixture are in progress. A cumbersome grid-search approach is also implemented, sometimes useful in special cases. McMaille is distributed under the GNU Public Licence conditions. The continuous computer speed increase will certainly favour further developments. A press-button mode is proposed to beginners, requiring the furniture of very few data for a default test.

[1] A. Le Bail (2002). McMaille program, http://www.cristal.org/McMaille/

[2] B.M. Kariuki, S.A. Belmonte, M.I. McMahon, R.L. Johnston, K.D.M. Harris and R.J. Nelmes (1999). *J. Synchrotron Rad*. **6**, 87-92.