

Limits of powder indexing of impure samples using whole profile methods

A. Le Bail

Université du Maine

Laboratoire des Fluorures

CNRS – UMR 6010



Content of the talk

- Introduction
- Starting and discussing some McMaille (pronounce « MacMy ») demos
and if time enough (certainly not)...
- Definition of the impurity
- « Simplicity » of the Monte Carlo algorithm in McMaille
- McMaille v2 : Gaussian profiles
- McMaille v3 : Columnar peak shape
- How many impurity lines can be tolerated ?
- Beyond impurity, indexing a 2-phases pattern, is that possible ?
- Conclusions



Introduction

Most powder pattern indexing software see their efficiency decrease a lot in presence of impurity lines.

Whole profile indexing computer programs were suggested to be less affected by this problem. But such programs are scarcely in the public domain.

Let us see how McMaille can perform. Let us even try to index simultaneously a two-major-phases powder pattern.

McMaille uses peak positions and intensities, then build a pseudo powder pattern with columnar peak shapes and test cell parameters by the Monte Carlo method up to find the best fit.

Starting and discussing some McMaille demos

<http://www.cristal.org/McMaille/>



Goal :

- 1 - Make a test : peak hunting with WinPLOTR + Indexing by McMaille v3 in automated mode.
- 2 - Indexing in manual mode – same powder pattern
 - a – adding 6 impurity lines to 20 lines (23%) – total 26 lines
 - b – adding 12 impurity lines to 20 lines (37.5%) – total 32 lines
 - c – more difficult : 10 lines of 20 are impurity lines (50%)
- 3 – Indexing a (simple) two-phases (both cubic) pattern (30 lines : 2x15)
- 4 – Seeing the result of a more complex two-phases case (tetragonal + orthorhombic) needing more than 20 minutes of calculations.

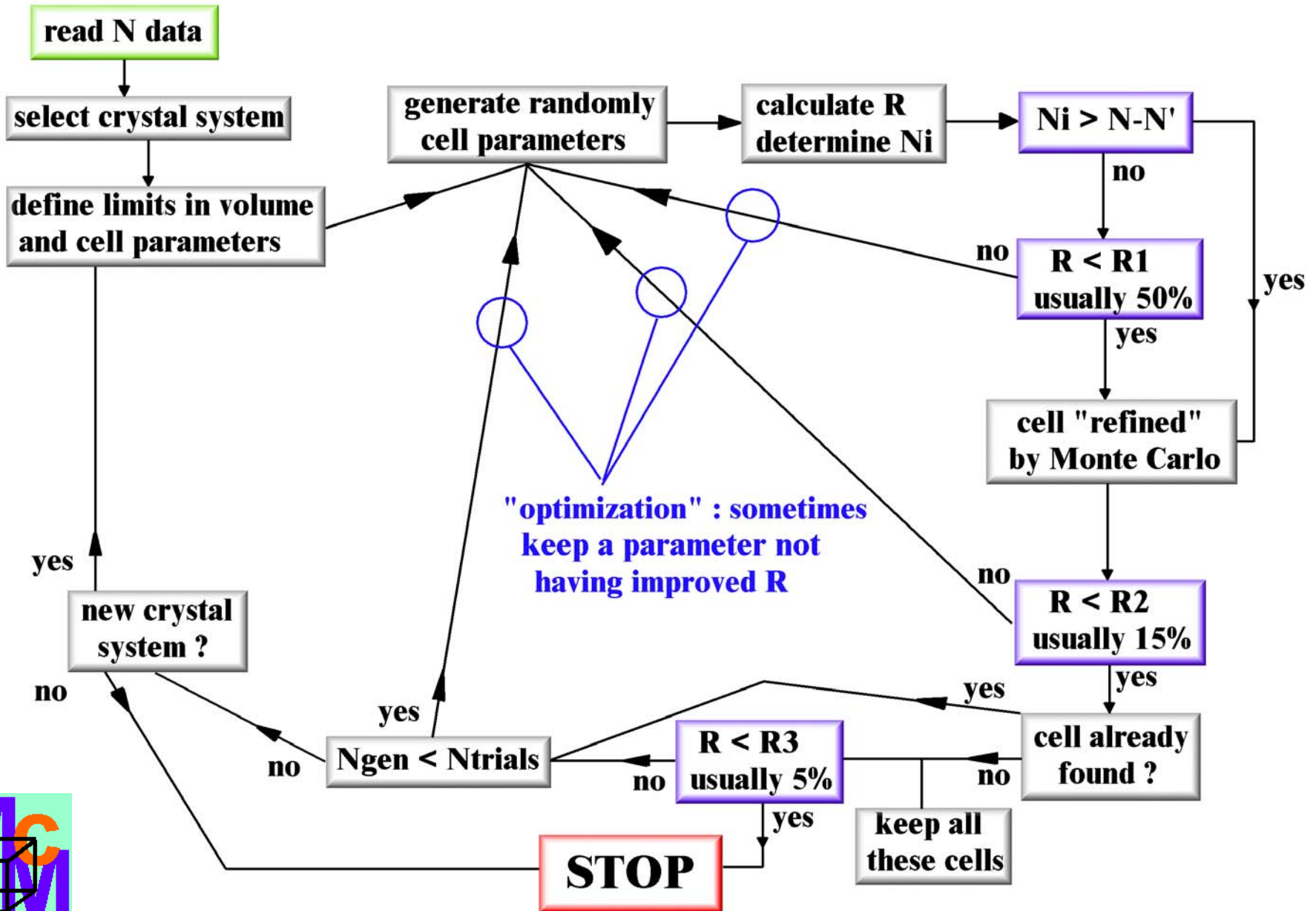
Definition of the impurity

Obviously, it should represent less than some percentage of the total intensity in the powder pattern.

Arbitrarily, let us take $I(\text{impurity}) < 15 \% I(\text{total})$

Clear that a software claiming at some insensitivity about impurities should take account of peak intensities : this is the case of profile indexing software like McMaille.

« Simplicity » of the Monte Carlo algorithm in McMaille



McMaille v1 and v2 : Gaussian profiles

- Choice of an idealized profile (Gaussian shape applied to extracted peak positions) rather than using the raw pattern – for velocity reasons. Fit by 3 iterations of the Rietveld decomposition formula (= Le Bail method).

-**Version 1** worked only in cubic for studying the feasibility which was quite encouraging with 1000 tests per second by using a > 2GHz processor.

-**Version 2** extended to all crystal systems, 300 tests per second in triclinic.

**Not fast enough with low symmetries needing
 10^8 - 10^9 tests...**



McMaille v3 : Columnar peak shape

Speed increased by a factor 20...

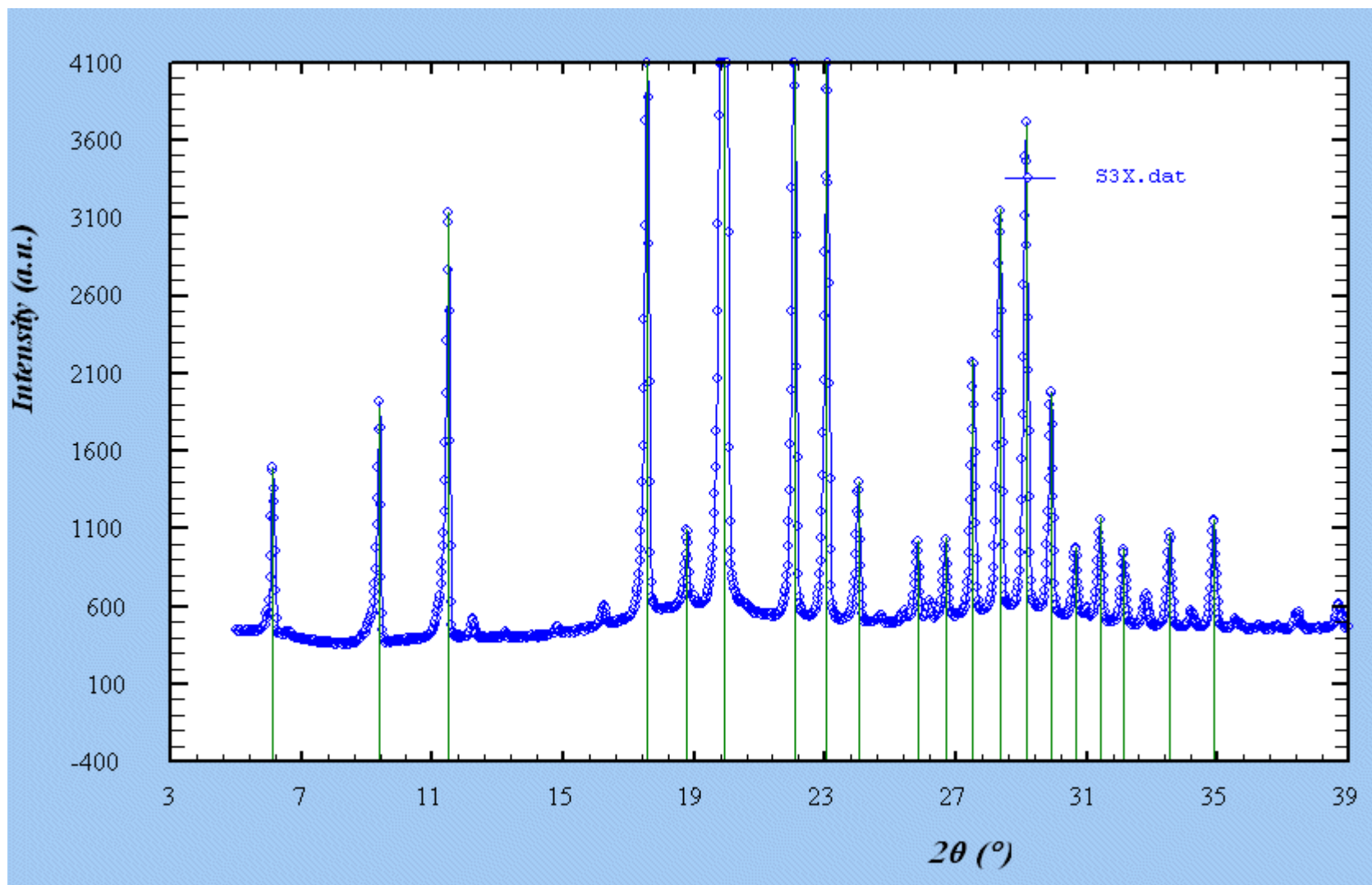
20.000 tests per second in cubic, 6000 in triclinic (for 20 peaks and no unindexed line...).

No real need for a fit, the observed and calculated columns are given the same height and same width.

The R factor becomes function of the percentage of overlapping between observed and calculated columns.

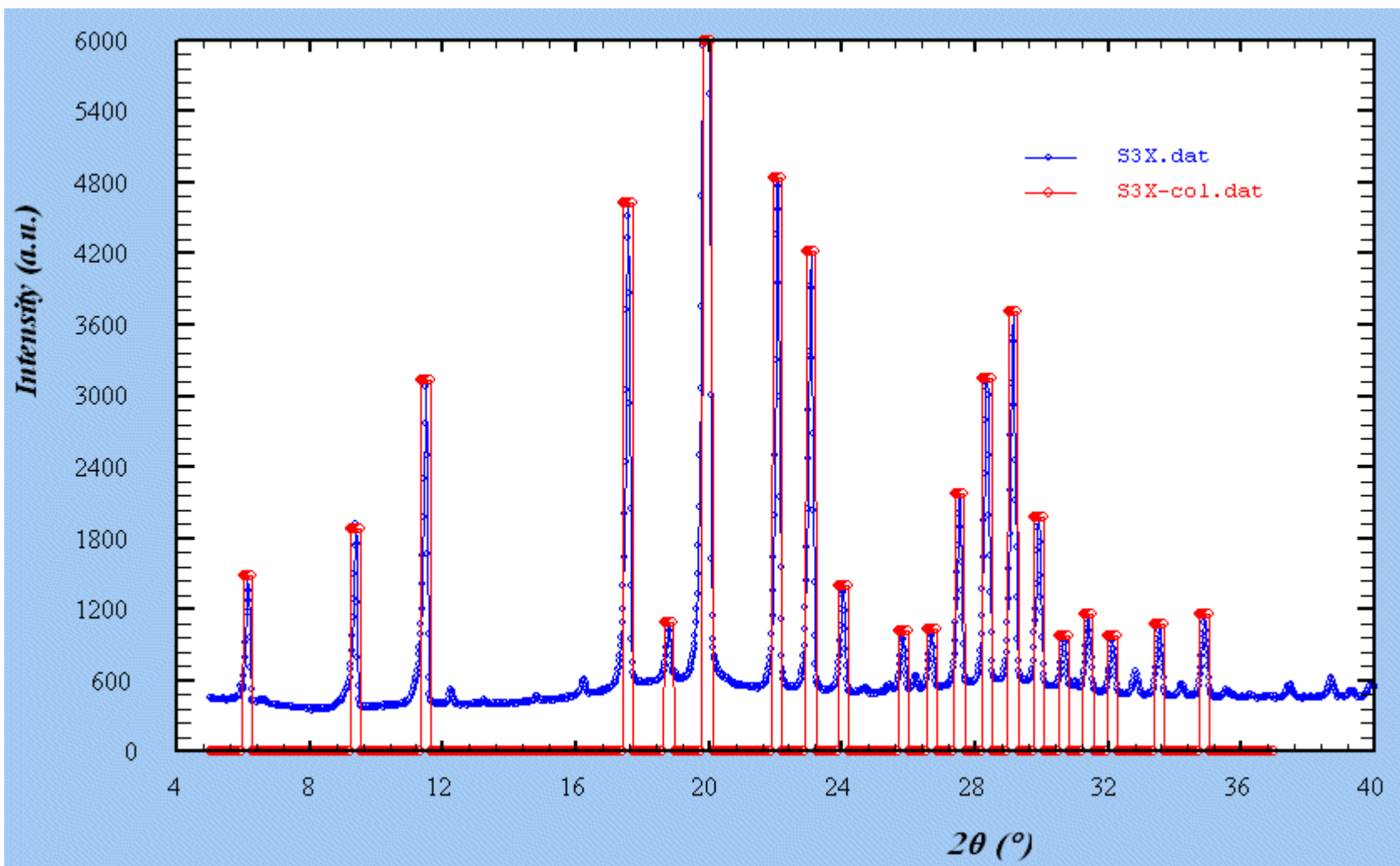


SDPDRR2 Sample 3 – conventional X-rays

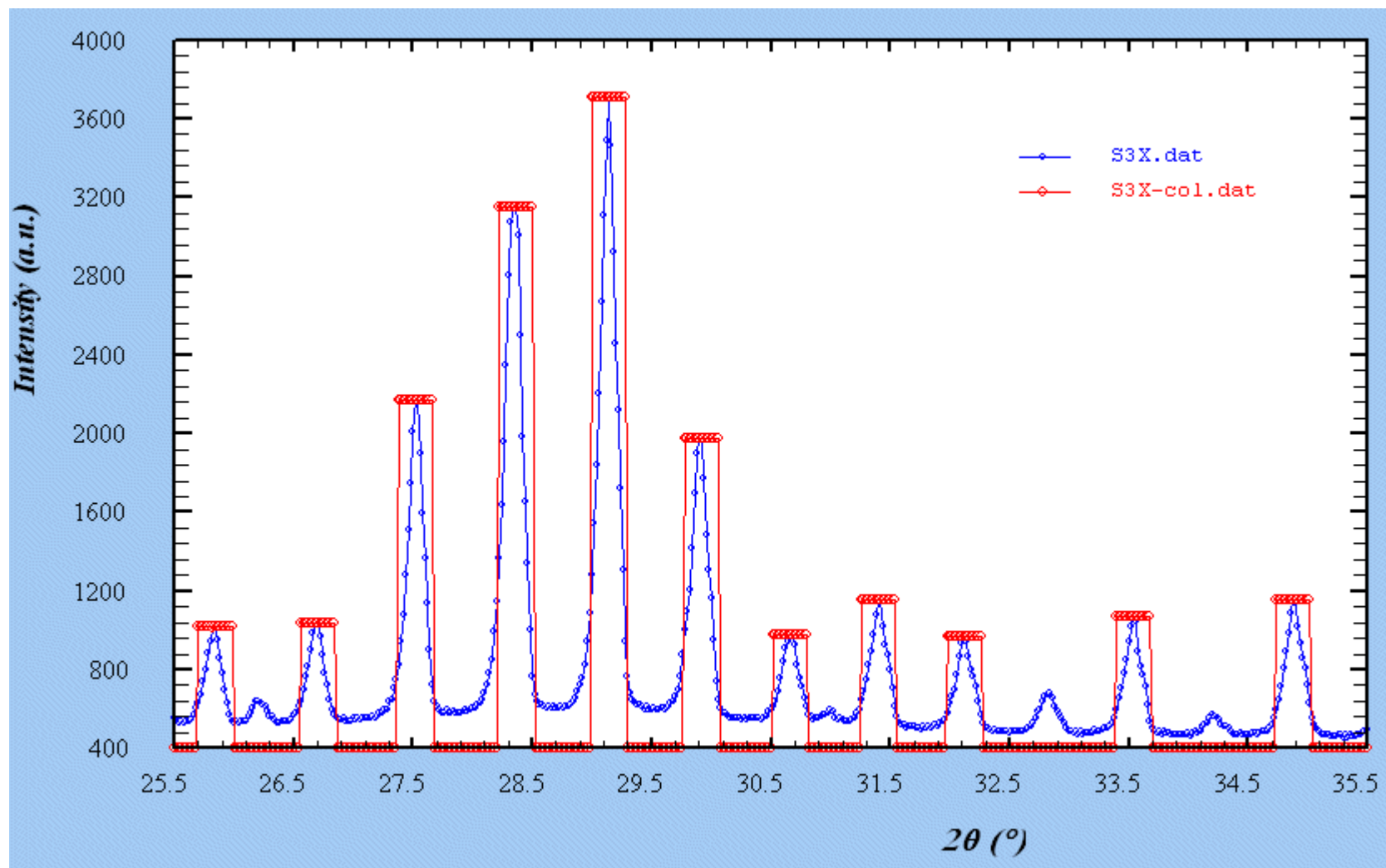


Peak positions extracted by WinPLOTR

The columnar peak shapes used by McMaille v3



Zooming on the last reflections

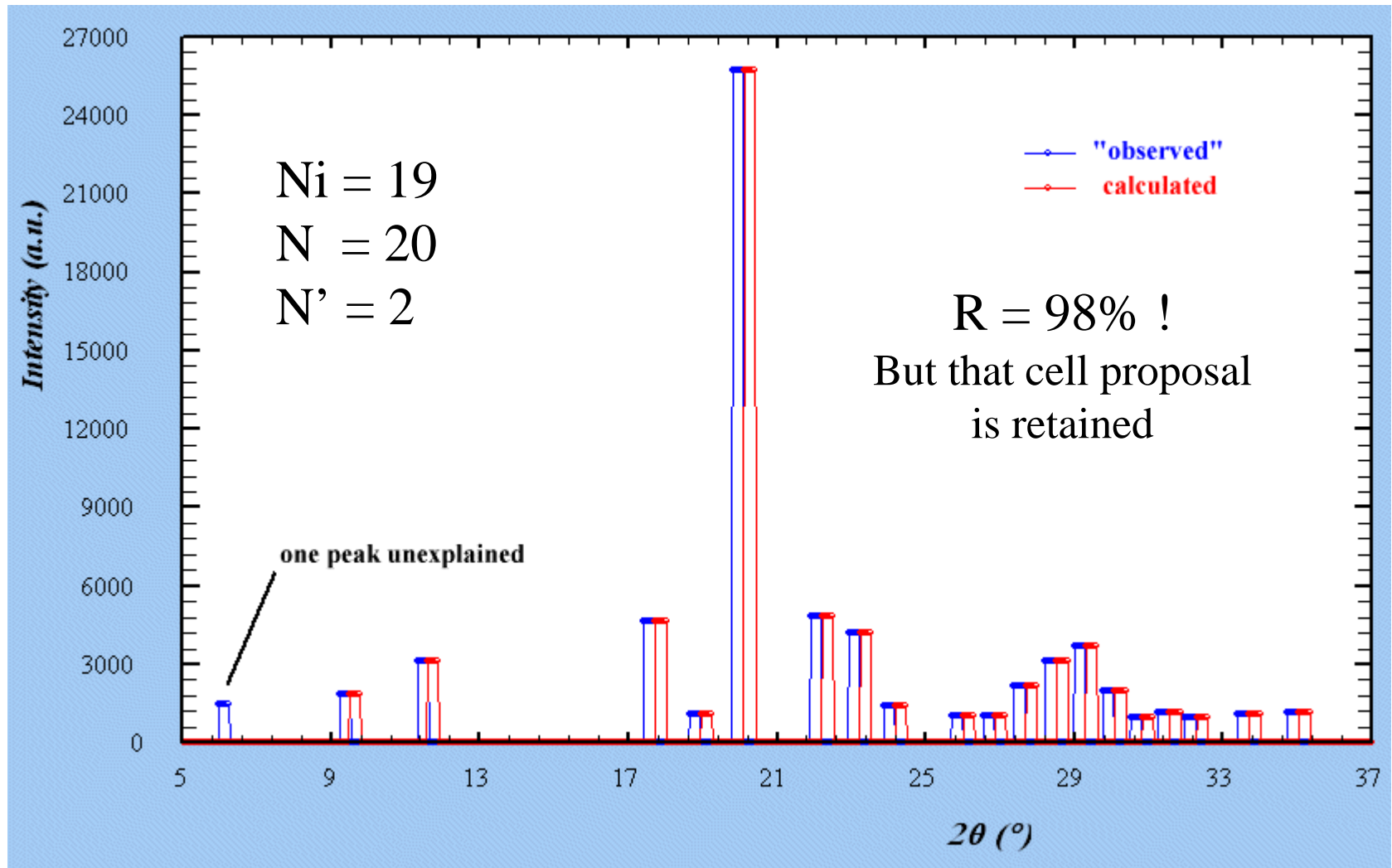


The width in the automated mode is calculated as :

$$0.3 \times \lambda / 1.54056$$

it depends on the user in manual mode

One of the 2 cases leading to « refine » a cell : $N_i > N - N'$



A peak is considered as indexed if some overlap occurs with a calculated one.
The second case is when $R < R_1$ (usually 50%).

About (accuracy + enlarged peaks) apparent contradiction

The more the « observed » columns are large, the more you have chances to intercept them by the calculated columns.

Cubic example : A column at $10^\circ(2\theta)$ ($d = 8.838 \text{ \AA}$) will extend from 9.85 to $10.15^\circ(2\theta)$ ($d = 8.972$ to $d = 8.707$; $\lambda = 1.54056 \text{ \AA}$). So that if the peak is the 200 reflection, the range of a values leading certainly (if the accuracy is high) to the solution is $[17.41-17.94 \text{ \AA}]$.

Any test in that interval larger than 0.5 \AA is a winning test.

But at the « refinement » stage, it is mainly the position accuracy which is important. It will lead effectively to low R values (exact overlapping corresponds to $R = 0$.) allowing to distinguish the true solution from bad proposals.



More on the cell « refinement » in McMaille

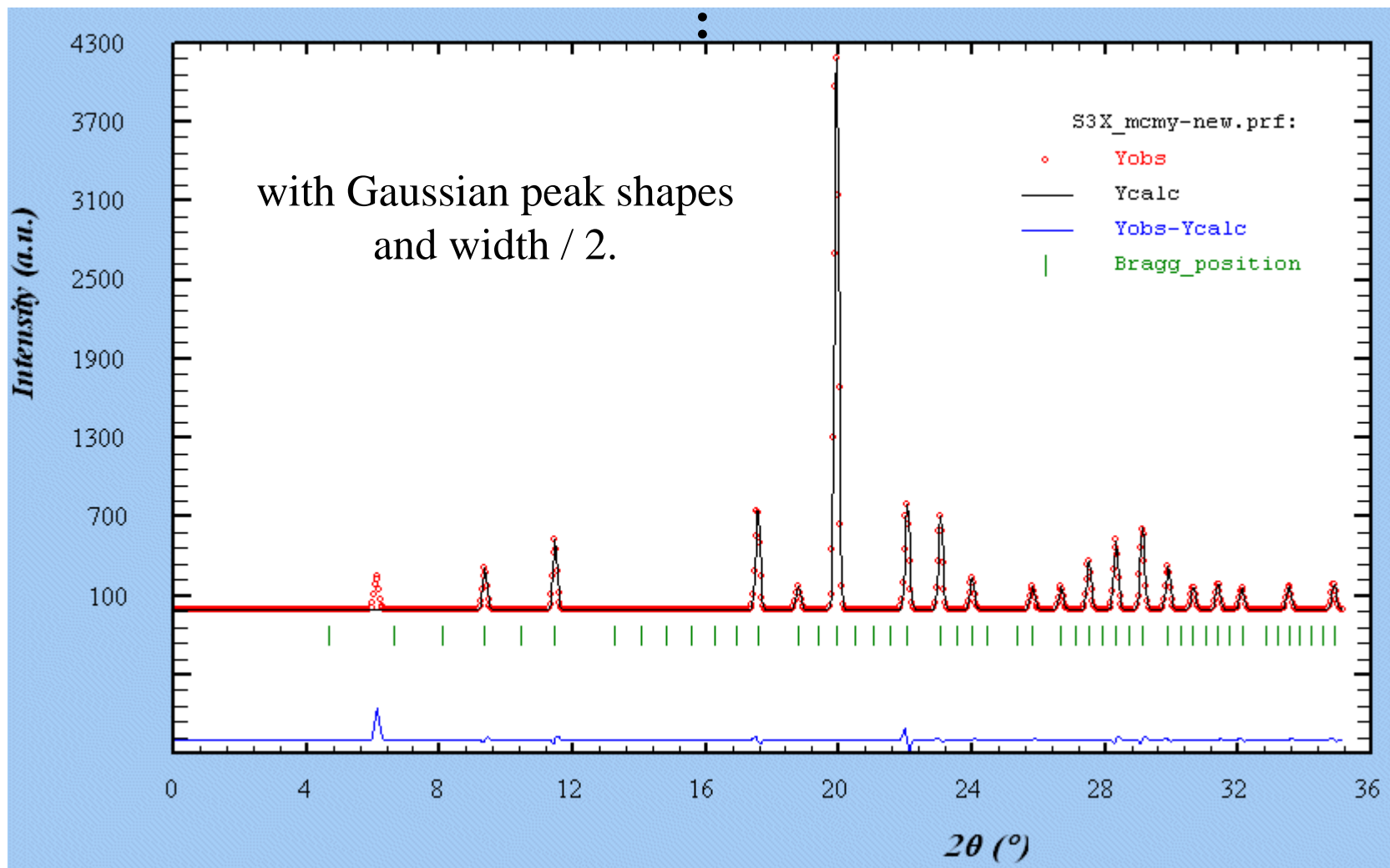
At $R < R1 \sim 50\%$, no least square refinement is possible

So that the cell parameters are adjusted by Monte Carlo (200 steps in cubic to 5000 steps in triclinic) with small amplitude change ($|\delta \max| = 0.02 \text{ \AA}$). Similar to tempering ?

The best proposal at $R < R3 \sim 5\%$ is finally least-squared refined and classical figures of merit (M_{20} , F_{20}) are calculated.



Final plot produced by McMaille, displayed by WinPLOTR



Other software compatible with McMaille outputs :
CHEKCELL, CRYSFIRE

Optimization : not being trapped in a false minima

Accepting a parameter change even if the fit is not improved

Effects with various **Probability** values (probability to accept a new cell parameter if the fit is not improved) : number of times the correct answer is found for the same number of Monte Carlo steps :

P (%)	0	15	30	45	60	75	100
Test 1 – orthorhombic	41	45	32	27	15	6	1
Test 2 – rhombohedral	28	41	40	28	17	10	6
Test 4 – monoclinic	47	60	46	45	25	19	2
Test 6 – triclinic	36	42	36	24	18	12	12

The tendency is to work better with P ~ 15 %, as a mean

P : a value of 15 means that in 15% of the tests, a parameter change may be accepted even if that change does not lead to any R decrease or number of indexed reflections improvement (no change means that you keep the previous parameter unchanged)

P = 100 : always accepted even if it does not improve the fit

P = 0 : not accepted at all if it does not improve the fit



Relative insensitivity to impurity

The user decides by two control parameters :

N' : number of unindexed lines.

R2 : consider only proposals with $R < R2$. Fixing it at 15% means that cell proposals explaining 85% of the peaks total intensity will be listed.

An impurity should not concern more of 15% of the total intensity, right ? But the number of (small) peaks belonging to the impurity can be high...





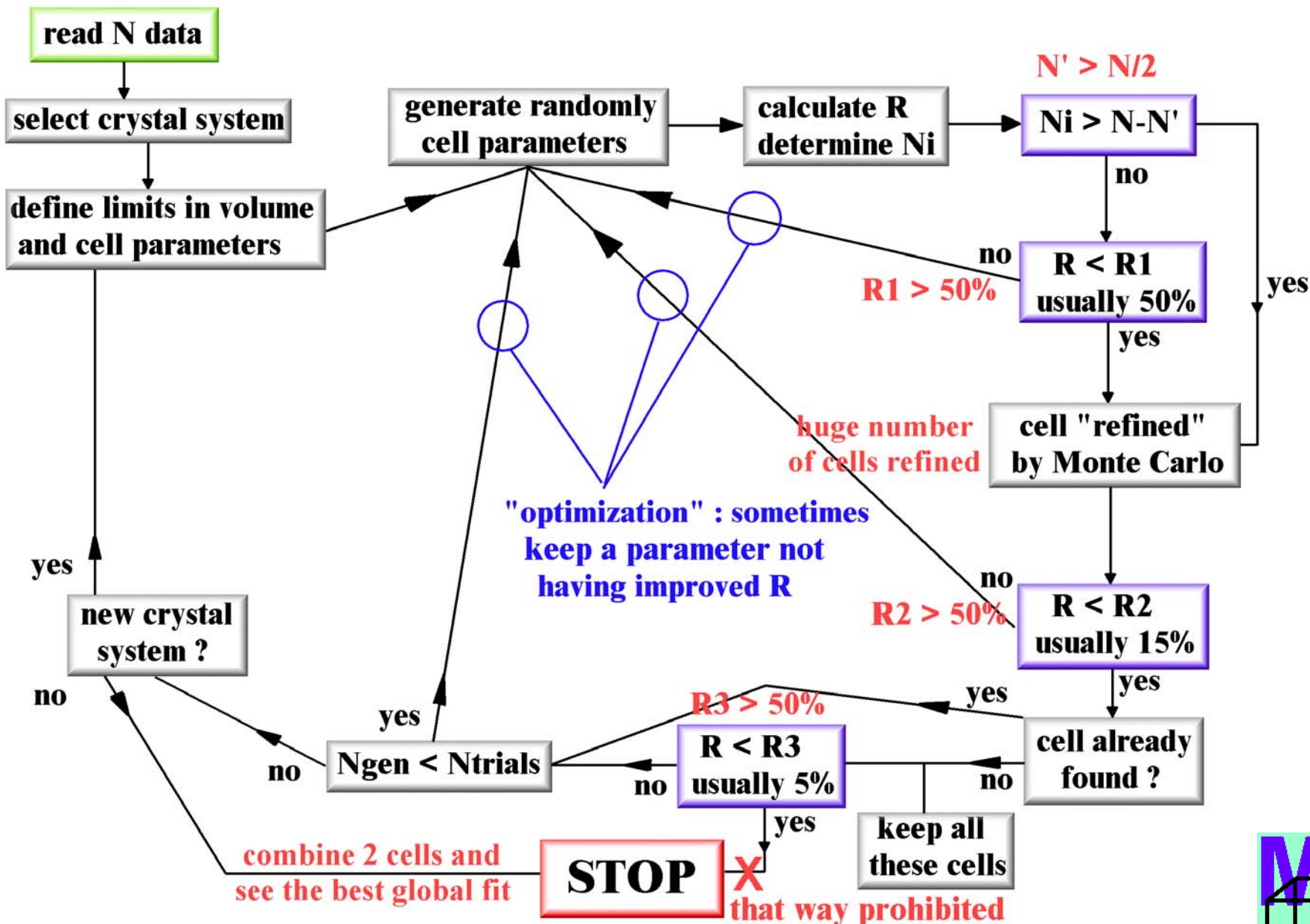
Beyond impurities, indexing multiphase patterns

Multiple synthesis in varying conditions should reveal the multiphase nature of the sample.

It is much better to adjust the synthesis conditions, and even if the phases cannot be prepared as pure phases, intensities variations should allow to define the peaks belonging to one or the other phase.

But if really you want to attempt indexing of a mixture, let us see the cost on the McMaille organization chart...

Indexing a 2-phases powder pattern with McMaille



Results of the McMaille demos

Conclusions from the demos are (not?) very encouraging :

With 50% (in number) of impurity lines (but representing less than 15% in intensity), the correct solution is hard (if not impossible) to distinguish from the bad ones.

With 6/26 (23%) or 12/32 (37.5%) of impurity lines, the correct solution may appear in first position (with the lowest R). But will you believe to it since the figures of merit M_{20} and F_{20} become much smaller ?

Attempting two-phases indexing in all crystal systems is not possible (with McMaille) before considerable increase in computer speed. You may have some success with symmetries equal or higher than orthorhombic. The correct answer is not always in first position, so will you have the courage to test all proposals ? Even if you succeed in the indexing, considerable effort may be needed then to solve two structures simultaneously.

About three-phases indexing or more : forget it ! Make some progress in chemistry...

General Conclusions :

Using the raw pattern is for tomorrow (too slow) but using a pseudo-pattern is for today (fast enough)

Indexing with care of peak intensity is the correct way for less sensitivity to impurities – however, the more impurity lines, the more difficult is to distinguish the correct cell among the bad ones.

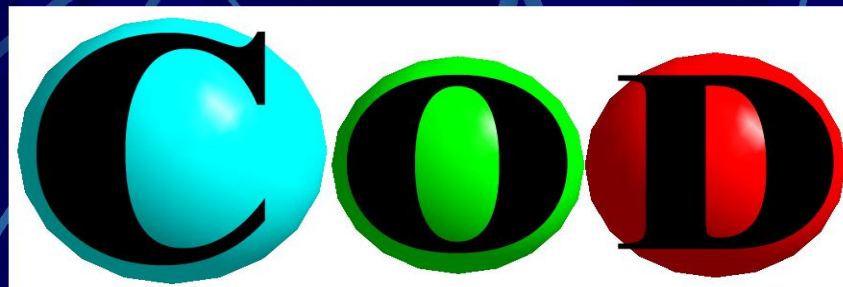
Use cautiously the 2-phases mode with McMaille – it is time-consuming and not always satisfying...

A useful address for a distance learning course :

SDPD Internet Course

<http://sdpd.univ-lemans.fr/course/>

Crystallographers join the Crystallography Open Database



www.crystallography.net



Deposit your crystal data
in the Public Domain
Thanks !



CRYSTAL
IMPACT



Advisory Board :

Michael Berndt, Daniel Chateigner, Xiaolong Chen, Marco Ciriotti, Lachlan M.D. Cranswick,
Robert T. Downs, Armel Le Bail, Luca Lutterotti, Hareesh Rajan, Alexandre F.T. Yokochi