

For the filters realized in analogue circuitry the preliminary filtration is preferable to carry out with the help of Chebyshev filters which provide acceptable non-uniformity of group delay time from the point of view of acoustical perception of admissible distortions and provide rather big attenuation on boundary frequency of a leakless strip.

For receiving the linear phase-frequency characteristics in ADC and DAC with redigitization it is necessary to use non-recursive digital filters.

## Bibliography

1. Vologdin, E. How can arise and sound the quantization / E. Vologdin // Audio operator. 2006. P. 28–41; 2007. P. 32–40. (in Russian)
2. Obolonin, I. A. The estimation of influence of analogue-digital conversion on distortions at a compression of digital audio data / I. A. Obolonin // The Bulletin of the Siberian State university of telecommunications and informations : proceedings. Novosibirsk, 2008. № 2. P. 67–71. (in Russian)

© Obolonin I. A., Rygovskaya N. A., 2009

Ya. I. Yakimov

Siberian Federal University, Russia, Krasnoyarsk

## TWO-LEVEL GENETIC ALGORITHM FOR X-RAY POWDER DIFFRACTION STRUCTURE ANALYSIS

*A new evolutionary approach for crystal structure determination of powders based on X-ray diffraction full-profile analysis and genetic algorithm of global optimization is suggested. An investigation of efficiency of given algorithm is carried out on test real-world problems of structure determination.*

*Keywords: evolutionary algorithm, X-ray powder diffraction analysis, Rietveld method.*

Crystal structure information is essential for explanation and prediction of physical and chemical properties of investigated materials. Many materials, multi-phase mixtures in particular, are available in form of powder only, thus severely impeding a research. In such cases X-ray powder diffraction methods, which are being intensively developed during last two decades, are used. They are based on analysis of a whole X-ray diffraction profile of powder pattern, which is a monochromatic X-ray radiation intensity function of polycrystalline sample diffraction angle. By now, in general, crystal cell parameters search problem (indexing methods) and structure model refinement problem (Rietveld method) have been solved. Primary mathematical means used for these problem solutions is a non-linear least-squares method (LSM). Plausible structure model determination in case of powder samples is still a problem even in case of relatively simple structures [1].

In recent years, for this problem solving so-called “direct-space” methods [1] have become of use. They are based on probabilistic generation of trial crystal structure models, their assessment through weighted difference of calculated and observed patterns (profile R-factor) and search for global minimum over corresponding parametric hypersurface in order to find an adequate structure model. An example of this approach is evolutionary algorithm, mimicking processes of natural selection in search of an optimal structure solution [2]. Several implementations of this concept have already been used for structure determination, demonstrating promising prospects [1; 3]. Here a two-level hybrid genetic algorithm is suggested for that purpose and its approval results are described on real patterns of single- and multi-phase polycrystalline samples with well-known crystal structure.

**Full-profile crystal structure model refinement.** As a tool for crystal structure model refining multi-phase Rietveld method [4] was used. An essence of Rietveld method is a modeling of experimental pattern by complex multi-parametric function:

$$Y^{\text{mod}}(\mathbf{P}, \theta_j) = \sum_i^n \Omega_i(\mathbf{P}_L, \theta_j) * I_i^{\text{calc}}(\mathbf{P}_S, \theta_j) + B(\mathbf{P}_B, \theta_j), \quad (1)$$

where  $\theta_j$  – diffraction angle;  $\Omega_i$  – profile functions for diffraction lines  $i$ , dependent on profile parameters set  $\mathbf{P}_L$ , (positions, half-width, form, asymmetry of lines, etc.);  $I_i^{\text{calc}}$  – calculated integral intensities of lines, dependent on structure parameters set  $\mathbf{P}_S$  (atomic coordinates, thermal motion parameters, etc.);  $B$  – function for background, dependent on background profile parameters set  $\mathbf{P}_B$ .

Firstly, pattern model is calculated from the approximate (initial) values of parameters  $\mathbf{P}$ , including model atomic coordinates in crystal cell. Exact coordinates and other parameters (including quantitative phase composition in case of multi-phase sample) are determined as a result of mathematical fitting of model pattern to observed pattern by structure and profile parameters least-squares method variation.

Formalizing the approach, we get a following mathematical optimization problem. Experimental data (powder pattern) represent a discrete sequence  $\{\theta_j, Y_j\}$  of size  $m$ , sorted by ascending of  $\theta_j$ . Some class of parametric function  $Y(\mathbf{P}, \theta)$  (Rietveld method functions) is given,  $\mathbf{P}$  is a set of profile and structural parameters (a vector of size  $N$ ),  $\theta$  – independent argument. Peculiarities of the problem are large dimensionality (can exceed 100 parameters) and non-polynomiality of functions.

With a given distribution of observed values of function  $Y^{\text{obs}}(\boldsymbol{\theta}) = \{\theta_j, Y_j\}$  and initial parameters approximation  $\mathbf{P}_0$  the task is to find function  $Y^{\text{mod}}$  of class  $Y(\mathbf{P}, \boldsymbol{\theta})$  and an optimal set of parameters  $\mathbf{P}^*$  to satisfy condition (2).

LSM functional:

$$\begin{aligned} \Phi(\mathbf{P}) &= \sum_j (Y^{\text{mod}}(\mathbf{P}, \theta_j) - Y^{\text{obs}}(\theta_j))^2 = \\ &= \sum_j (Y(\mathbf{P}, \theta_j) - Y_j)^2 \rightarrow \min_{\mathbf{P}}. \end{aligned} \quad (2)$$

As a figure-of-merit of LSM solution, according to [4], weighted difference of calculated and observed patterns (profile R-factor) is taken:

$$R = \sqrt{\frac{\sum_j (Y^{\text{mod}}(\mathbf{P}, \theta_j) - Y^{\text{obs}}(\theta_j))^2}{\sum_j (Y^{\text{obs}}(\theta_j))^2}} \cdot 100 \%. \quad (3)$$

A necessary condition of extremum for (2):

$$\begin{aligned} \sum_j (Y_j - Y(\mathbf{P}, \theta_j)) \frac{\partial Y(\mathbf{P}, \theta_j)}{\partial P_k} &= 0, \\ k &= 1, \dots, N, \end{aligned} \quad (4)$$

where  $P_k$  –  $k$ -th component of parameter vector  $\mathbf{P}$ .

However, due to non-linearity of functions  $Y(\mathbf{P}, \boldsymbol{\theta})$  over parameters  $\mathbf{P}$  derived system (4) cannot be solved analytically. Linearizing the system (4) by Taylor expansion at starting point  $\mathbf{P}_0$  with truncating terms above first order, we can obtain a system of linear equations over  $N$  variables (non-linear least-squares method). Iteratively solving for  $\Delta \mathbf{P}_1, \Delta \mathbf{P}_2, \dots$  (refining values  $\mathbf{P}: \mathbf{P}_1 = \mathbf{P}_0 + \Delta \mathbf{P}_1, \dots$ ), we will move towards  $\mathbf{P}^*$ . Solution process convergence is defined by proximity of point  $\mathbf{P}_0$  to optimal  $\mathbf{P}^*$ . With starting point  $\mathbf{P}_0$  declining and problem dimensionality  $N$  increasing the iterative method becomes unstable and starts to diverge. The bigger  $N$  (or the worse the quality of experimental data), the more precise  $\mathbf{P}_0$  are required, which practical determination represents a serious obstacle.

**Genetic algorithm of structure analysis.** Effectiveness of evolutionary algorithms in complex non-linear global optimization problem solving was proven [2]. So the idea arose to combine the Least-squares method of seeking the minimum of functional (2) with evolutionary algorithm of objective function (3) optimization in order to solve above-stated problem. A two-level evolutionary algorithm comprising two distinctive genetic algorithms (GA) is suggested.

**First level GA.** The first level of proposed algorithm is a «conventional» hybrid GA [2; 3], dealing with binary representation of parameter values. Its chromosomes encode vector of sought parameters  $P$ , where binary representation accuracy is varied by user. In particular, fragments of chromosome define rounded with specified accuracy coordinates  $x, y, z$  of atoms of investigated material relatively to its elementary unit cell. Minimized objective function over  $P$  is R-factor (3), ideally tending to zero while converging toward a global minimum. R-factor is calculated unambiguously for each parameter vector  $P$  with given sampling  $Y^{\text{obs}}(\boldsymbol{\theta})$  and in practice, depending on simulation and experimental error magnitudes, should come to about 5...10% (defined empirically) in optimal point.

A flow-chart of first-level algorithm is shown in figure 1 on the left. Starting population is generated arbitrarily, so a priori given starting approximations are not required. Tournament selection of parents with varied tournament size is employed. Algorithm, besides standard genetic operators – recombination (1 – point, 2 – point or uniform) and mutation, uses local search operator. The best individual and some randomly chosen individuals are subjected to LSM local descent over all coordinates (modified Newton–Raphson algorithm). Lamarckian concept of evolution [2; 3] is implemented, where parameter values found by the local search replace old ones.

Primary objective for this level is to find plausible (in the sense of R-factor) initial approximations of parameters  $\mathbf{P}_0$ .

**Second level GA.** Evolutionary algorithm of the second level is utilized for the searching and generating of LSM refinement strategy for initial approximations of parameters  $\mathbf{P}_0$ , representing a sequence of local descents on R-factor hypersurface. Bit strings  $\mathbf{B}_i$  defining groups of parameters refined on current generation are used as this level individuals. The length of a bit string equals to a number of sought parameters  $N$  with each bit corresponding to certain parameter. Bit value on a position  $k$  of the second individual indicates whether to refine (= 1) or not (= 0) the  $k$ -th parameter on a current iteration. The values of parameters  $\mathbf{P}$  for the every string are refined iteratively with non-linear LSM from (4). For example, string 101 means that equations (4) are constructed only for  $k = 1$  and  $k = 3$  and after solving of  $2 \times 2$  system give increments for № 1 and 3 parameters, while parameter of index 2 is left unmodified. Thus, each  $\mathbf{B}_i$  defines a search sub-space.

A flow-chart of second-level algorithm is given in figure 1 on the right. Initially this level GA individuals can be generated arbitrarily or according to some empiric scheme based on user-provided patterns sequence (masks imposed on  $\mathbf{B}_i$ ). For assessment of the GA individuals for each  $\mathbf{B}_i$  a relation to first-level  $\mathbf{P}_i$  individual (one or many) is set. For  $(\mathbf{B}_i, \mathbf{P}_i)$  pairs the LSM is applied according to above principles, with a result of  $\mathbf{P}'_i$  – refined  $\mathbf{P}_i$  values in accordance with  $\mathbf{B}_i$  string. Level 2 objective function takes into account a performance of applied LSM: as figure-of-merit (fitness) of the 2nd level individuals the function (5) is taken:

$$F = [R(\mathbf{P}'_i) / R(\mathbf{P}_i) + p]^{-1} \quad (5)$$

where  $p$  – a penalty for non-convergence (substantial increase of local search steps lengths).

Thus, the better the refinement process convergence, the higher the assessed individual fitness.  $\mathbf{B}$  – individuals are recombined and mutated without  $\mathbf{P}$  – individuals altering. Results of that evolution are the refinement strategies of  $\mathbf{P}$  – individuals.

Objective for this level is to carry out a sequence of refinements (local searches over dimensions specified by bit string) using the best solutions from the preceding level. Providing sufficiently suitable initial approximations  $\mathbf{P}_0$ , the refined  $\mathbf{P}$  will converge to optimum. Unsuitable  $\mathbf{P}_0$  with many refined parameters will not yield convergence; in this case, a subsequent executing of first level algorithm can give better initial approximations for second level. The best parameter strings  $\mathbf{P}$  are returned to the first GA level for assessment and inclusion in next population  $\{\mathbf{P}_i\}$ .

The proposed algorithm involves a cyclic executing of both levels while stop criteria are not satisfied (computational resource is exhausted,  $\min R < R_{\text{stop}}$ ,  $\text{mean } R < R'_{\text{stop}}$ ). One or many better and some randomly chosen individuals are transferred to another level. Evolution of trial structures on the first level provides a search for values suitable for second level minimization by an evolutionary sequence of local searches. Moreover, this approach affords overcoming of local minima, where LSM can fall to during second level performing. Thus, stochastic and deterministic search procedures are combined here mutually complementing and enhancing each other.

The algorithm has been implemented as a shell over a DDM console program incorporating Rietveld-like method [5]. For the purpose of the algorithm performance evaluation it was tested on single- and multi-phase samples with known crystal structure of component phases (samples 1, 2).

**Sample 1.** A determination of crystal structure of component phases and quantitative phase analysis of three-phase sample CPD-1h, given by Commission on Powder Diffraction of International Union of Crystallography at Round Robin on QPA [6] quantitative phase analysis contest.

Simultaneously, profile parameters, coordinates of atoms in general crystallographic positions and thermal atomic parameters (29 parameters in total) were searched through all possible value ranges. A convergence plot for two-level GA indicating the R-factor decrease during parametric and bit strings populations evolution is given in figure 2. It is

clear that R-factor decrease is primarily provided with 2nd GA level, while 1st level efficiently leads out from local minima. An optimal solution was obtained after 5th full GA cycle. It was empirically drawn that 3 generations on each GA level with population sizes of 20 1st level and 10 2nd level individuals are sufficient for reliable convergence of the method. Population sizes are comparable with the dimensionality of the problem that indicates an efficient usage of computational resource and substantial potential of the method. Time spent for the problem solving: 4 min 47 sec (CPU AMD X2 4400+).

On a final stage of the algorithm, the phase concentrations were calculated. A correspondence between true and found compositions serving as an integral quality criterion of obtained solution is shown in table 1. In the last table row the root mean square (RMS) is given.

It should be noted that RMS of phase analysis solutions from Round Robin on QPA for CPD-1 samples averages  $\sim 3\%$  mass [6].

**Sample 2.** Crystal structure of single-phase sample  $\text{Pd}(\text{NH}_3)_2(\text{NO}_2)_2$  [7] determination.

All coordinates of atoms in general crystallographic positions (including coordinates of hydrogen atoms) and thermal atomic parameters (26 parameters in total) were searched simultaneously. A corresponding GA convergence plot is demonstrated in figure 3. As in previous case, 3 generations of each level were generated on a full GA cycle. Used population sizes: 20 individuals on the 1st level and 10

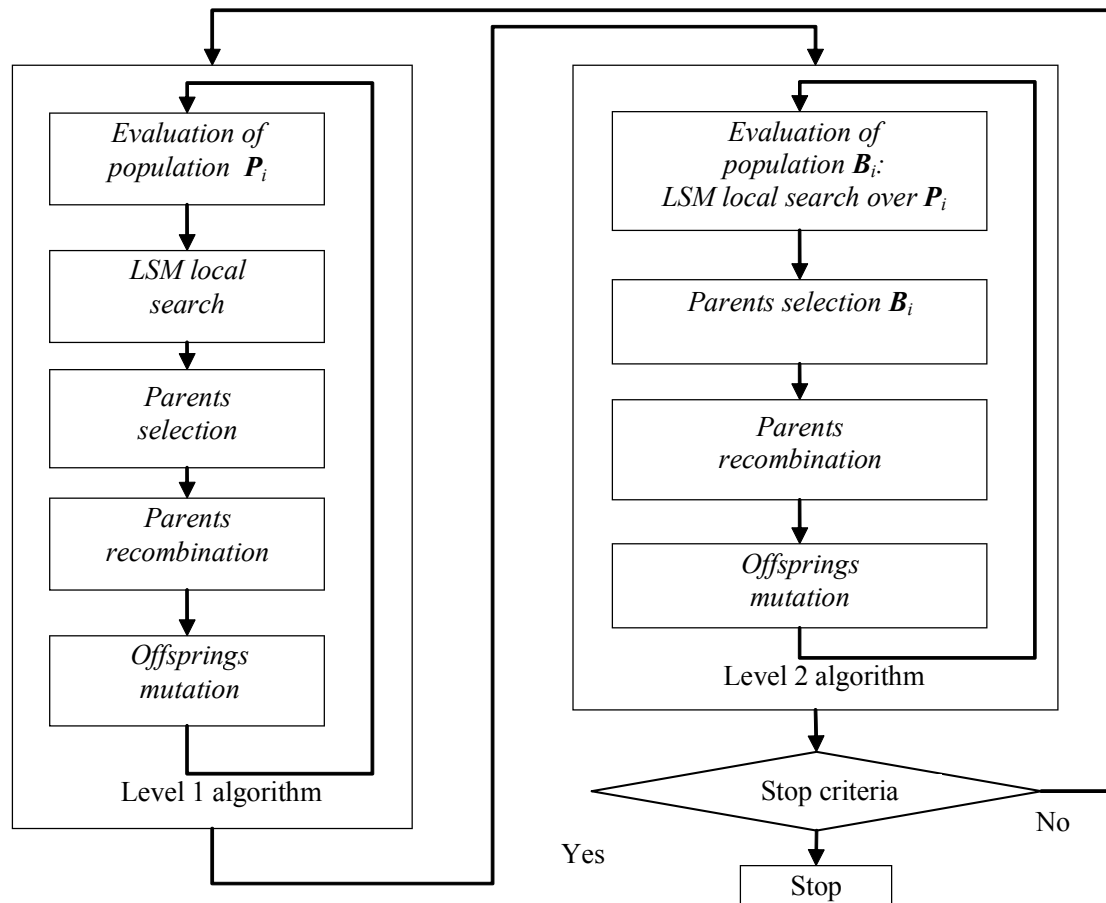


Fig. 1. Two-level GA flow-chart

on the 2nd level. An optimal solution was obtained after 3th full GA cycle. One can see that convergence here again is primarily provided by the 2nd GA level. Time spent: 4 min 21 sec (CPU AMD X2 4400+).

A correspondence between experimental and calculated X-ray powder patterns from the last GA stage is demonstrated in figure 4. It can be an integral quality criterion of obtained solution.

Found coordinates of the atoms (relative to the crystal cell axes) and thermal parameters compared to reference values [7] (designated with asterisks) are given in the table 2.

Obtained maximum error for coordinates of the heavier atoms: 0.0015, for thermal parameters: 0,012, and for coordinates of the hydrogen atoms: 0.0170.

The accuracy of obtained solution suits the accuracy of reference model structure [7]. It should be noted that with GA the coordinates of hydrogen atoms were found, which, being the lightest of all atoms, is hard to locate with available powder pattern analysis methods.

Described two-level GA comprises search and refinement of crystal structures thus giving the possibility of automation of structure determination process. Important features of

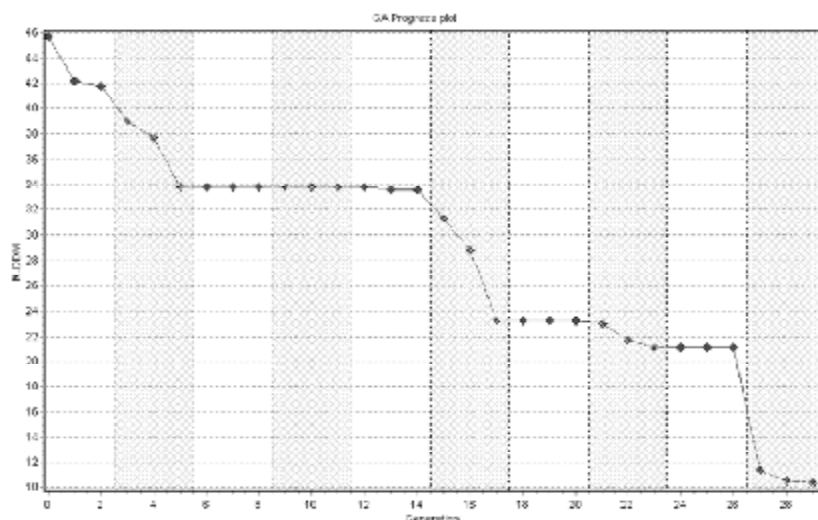


Fig. 2. GA convergence plot. The best found so far (to a current generation) solutions are designated (x-coordinate – the number of generation, y-coordinate – R-factor). Cross-hatching marks second-level GA executing

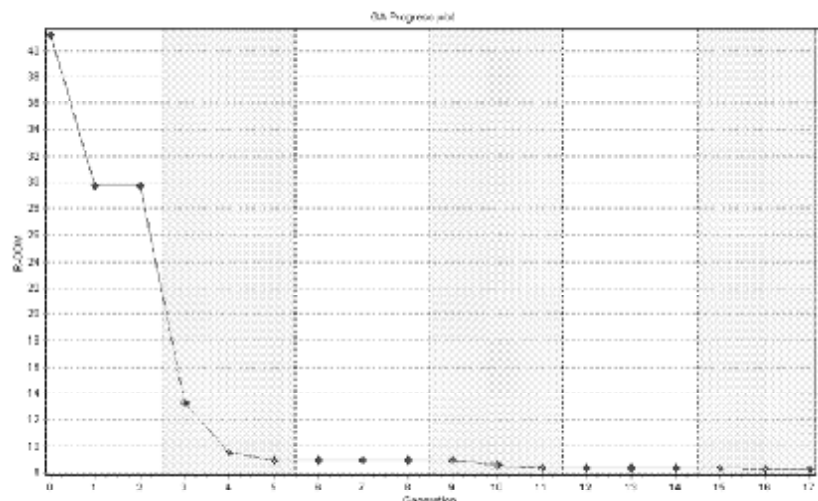


Fig. 3. GA convergence plot. The best found so far solutions are designated. Cross-hatching marks second-level GA executing

Table 1

CPD-1h sample composition

Phase	Formula	True (% mass)	Found (% mass)	Error (% mass)
Corundum	Al <sub>2</sub> O <sub>3</sub>	35.12	35.39	0.27
Fluorite	CaF <sub>2</sub>	34.69	35.08	0.39
Zincite	ZnO	30.19	29.53	0.66
RMS				0.57

algorithm are, as well, the possibility of simultaneous search of profile and structure parameters, and in case of multi-phase samples, phase composition calculation. It seems that the key role here was played by the combination of first and second level algorithms. Apparently, proposed approach has substantial potential for further development.

### Bibliography

1. David, W. I. F. Structure determination from powder diffraction data / W. I. F. David, K. Shankland // Acta Cryst. 2008. Vol. A64. P. 52–64.
2. Michalewicz, Z. Genetic Algorithms + Data Structures = Evolution Programs / Z. Michalewicz. Berlin : Springer-Verlag, 1996.
3. Implementation of Lamarckian concepts in a Genetic Algorithm for structure solution from powder diffraction data

/ G. W. Turner, E. Tedesco, K. D. M. Harris et al. // Chem. Phys. Lett. 2000. Vol. 321. P. 183–190.

4. Bish, D. L. Quantitative phase analysis using the Rietveld method / D. L. Bish, S. A. Howard // J. Appl. Cryst. 1988. Vol. 21. P. 86–91.

5. Solovyov, L. A. Full-profile refinement by derivative difference minimization / L. A. Solovyov // J. Appl. Cryst. 2004. Vol. 37. P. 743–749.

6. Outcomes of the International Union of Crystallography Commission on Powder Diffraction Round Robin on Quantitative Phase Analysis: samples 1a to 1h / I. C. Madsen, N. V. Y. Scarlett, L. M. D. Cranswick, T. Lwin // J. Appl. Cryst. 2001. Vol. 34. P. 409–426.

7. Crystal Structure of trans-[Pd(NH<sub>3</sub>)<sub>2</sub>(NO<sub>2</sub>)<sub>2</sub>]: X-ray Powder Diffraction Analysis / A. I. Blokhin, L. A. Solovyov, M. L. Blokhina et al. // Rus. J. Coord. Chem. 1996. Vol. 22. P. 185–189.

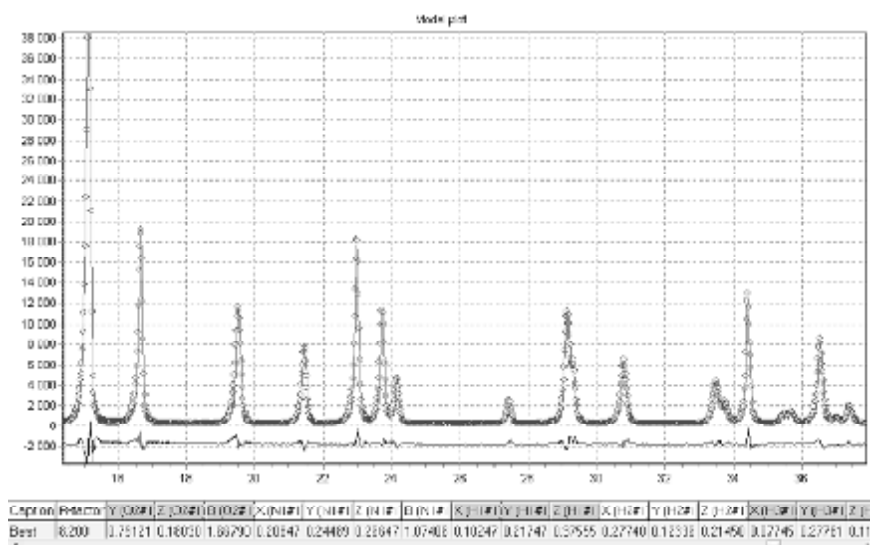


Fig. 4. Observed, model and difference X-ray powder patterns of the sample 2.

Observed data are designated with circles; difference curve is shifted down.

The model pattern is constructed using the best GA found values (given in the table at the bottom)

Table 2

### Crystal structure of Pd(NH<sub>3</sub>)<sub>2</sub>(NO<sub>2</sub>)<sub>2</sub> compound: reference and GA found

Pd(NH <sub>3</sub> ) <sub>2</sub> (NO <sub>2</sub> ) <sub>2</sub> . Space Group <i>P</i> – <i>I</i> (№ 2)												
Unit cell: <i>a</i> = 5.4251(1) Å. <i>b</i> = 6.3209(1) Å. <i>c</i> = 5.0031(1) Å.												
alpha = 111.87(0)°. beta = 100.4(0)°. gamma = 91.37(0)°												
At.	X*	X <sub>GA</sub>	\Delta	Y*	Y <sub>GA</sub>	\Delta	Z*	Z <sub>GA</sub>	\Delta	B*	B <sub>GA</sub>	\Delta
Pd	0.5000	–	–	0.5000	–	–	0.5000	–	–	0.484	0.485	0.001
N	0.3440	0.3448	0.0008	0.6970	0.6974	0.0004	0.3000	0.2999	0.0001	1.230	1.235	0.005
O1	0.1200	0.1205	0.0005	0.7270	0.7269	0.0001	0.2840	0.2835	0.0005	3.252	3.262	0.010
O2	0.4690	0.4688	0.0002	0.7910	0.7912	0.0002	0.1810	0.1803	0.0007	1.665	1.668	0.003
N1	0.2090	0.2085	0.0005	0.2450	0.2449	0.0001	0.2680	0.2665	0.0015	1.086	1.074	0.012
H1	0.1000	0.1025	0.0025	0.2220	0.2175	0.0045	0.3670	0.3756	0.0014	5.000	–	–
H2	0.2850	0.2774	0.0076	0.1260	0.1234	0.0026	0.2190	0.2145	0.0045	5.000	–	–
H3	0.0750	0.0775	0.0025	0.2770	0.2778	0.0008	0.0990	0.1160	0.0170	5.000	–	–

Note: Pd atom takes the special position in the center of cell, thermal parameters of the hydrogen atoms were fixed as they have insignificant impact on the calculations.