
Proceedings of the XXI International Meeting on Radio and Microwave Spectroscopy
RAMIS 2005, Poznań-Będlewo, Poland, April 24–28, 2005

Application of Genetic Algorithm for Extraction of the Parameters from Powder EPR Spectra

T. SPAŁEK^a, P. PIETRZYK^a AND Z. SOJKA^{a,b,*}

^aFaculty of Chemistry, Jagiellonian University
Ingardena 3, 30-060 Cracow, Poland

^bRegional Laboratory for Physico-Chemical Analyses and Structural Research
Ingardena 3, 30-060 Cracow, Poland

The application of the stochastic genetic algorithm in tandem with the deterministic Powell method to automated extraction of the magnetic parameters from powder EPR spectra was described. The efficiency and robustness of such hybrid approach were investigated as a function of the uncertainty range of the parameters, using simulated data sets. The discussed results demonstrate superior performance of the hybrid genetic algorithm in fitting of complex spectra in comparison to the common Monte Carlo method joint with the Powell refinement.

PACS numbers: 02.60.Pn, 02.60.Ed, 76.30.Fc, 82.75.Mj, 61.43.Gt

1. Introduction

Paramagnetic samples in the form of randomly oriented powders, glasses or frozen solutions give rise to the EPR spectra that usually are not easily amenable to direct quantitative analysis. In addition, low symmetry environments often encountered in disordered heterogeneous systems can further complicate the spectra [1]. Under such circumstances direct measurements based on the resonance line positions may give incorrect values, and advanced computer analysis and computer simulation are the only method suitable for accurately extracting the parameters [2, 3]. Single simulations, however, are of rather limited practical significance, because the number of parameters to be optimized may be large in many cases,

*corresponding author; e-mail: sojka@chemia.uj.edu.pl

and they encompass a wide range of possible values, which creates a strongly nonlinear multivariate optimization problem. Therefore, automated fitting is an indispensable component in any proficient analysis of the powder EPR spectra. The conventional local refinement methods used in the EPR spectra optimization strongly depend on the starting point, and their performance is rapidly hindered with the increasing number of parameters. Therefore, global techniques are employed to tackle optimization problems with many local optima. Among them the genetic algorithm (GA) based methods have recently received a great deal of interest, because of their outstanding performance [4, 5]. In this paper we describe an implementation of the genetic algorithm to optimization of EPR spectra and the evaluation of its performance in comparison to the Monte Carlo (MC) method, both joint with the Powell (PA) refinement, for automated analysis of complex calculated powder patterns such as might be observed for intrazeolite copper(II) complexes.

2. Spectra simulation and genetic algorithm premise

The test EPR spectra (*vide infra*) used in this study are parameterized in terms of the following spin Hamiltonian:

$$\mathcal{H} = \mu_B \mathbf{S} \cdot \mathbf{g} \cdot \mathbf{B} + \mathbf{I} \cdot \mathbf{A} \cdot \mathbf{S}. \quad (1)$$

In this formalism the \mathbf{g} tensor describes the electronic Zeeman interaction, the \mathbf{A} tensor — the hyperfine coupling between the electron spin $S = 1/2$ and the nuclear spin $I = 3/2$.

To simulate the spectrum, the set of adjustable spin Hamiltonian parameters along with the values of the intrinsic linewidths, $\mathbf{P} = \{\mathbf{g}, \mathbf{A}, \sigma(B) \dots\}$, has to be complemented by a collection of the associated parametric variables $\{\nu, S, I, \theta, \phi, h \dots\}$, containing auxiliary information about the resonance frequency ν , the values of S and I , the θ and ϕ angle increments used for integration as well as the number of the simulation points h . Once the full set of the parameters \mathbf{P} is known, the associated spectrum can readily be obtained by the simulation $\mathbf{Y}(B; \nu) = f(\mathbf{P})$ in the following way:

$$Y(B; \nu) = C \int_{\theta=0}^{\pi/2} \int_{\varphi=0}^{2\pi} \sum_i \mathbf{P}(i, \theta, \varphi, \nu) f(B - B_0[\nu], \sigma_B) d \cos \theta d \varphi, \quad (2)$$

where $\mathbf{P}(i, \theta, \varphi, \nu)$ is the transition probability, $f(B - B_0[\nu], \sigma_B)$ is the line shape function, and C is a constant that incorporates all instrumental parameters [6]. Actually the parameters can be extracted only through successful fitting of the simulated spectrum to the experimental one because the reverse procedure is unavailable. In praxis, due to exceedingly complicated mathematical nature of powder EPR spectra, the simulation process needs to be combined with advanced optimization methods to be efficient. In this context stochastic optimization techniques like genetic algorithm are particularly appealing since the spin Hamiltonian

search space \mathcal{H}_p is usually complex and with many local minima, so that conventional methods often fail to find the global solution [7].

For the optimization of multicomponent EPR spectra we have implemented a genetic algorithm (GA) combined with subsequent direct Powell (PA) refinement into a simulation program EPRsim32, as described elsewhere in more detail in [8]. The advantage of incorporating the local search method (PA) within the structure of the global genetic algorithm consists in synergistic amalgamation of deterministic and stochastic characters of both minimization procedures. To design the GA machinery, one should specify the size of an initial population $\Omega(m)$ of individuals \mathbf{G}_i , a coding method (representation), a fitness index $F(\mathbf{G}_i)$ of each individual, and the formulation of genetic operators of selection, recombination, and mutation.

In our implementation the full set \mathbf{P} of the spectral parameters was encoded into polychromosomic individuals $\mathbf{G}_i = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_c\}$, composed of c chromosomes \mathbf{X}_j , corresponding to the particular component signal $Y_j(B)$ of a total EPR spectrum. Each chromosome $\mathbf{X}_j = \{x_1^j, x_2^j, \dots, x_n^j\}$ is a string of genes x_n^j coded as 32-bit floating point numbers from the $[-1, 1]$ range. The number of genes is equal to the number of parameters n in the component signal j . The information contained in the chromosome \mathbf{X}_j can be converted from genetic space into the parameter space using the following translation function:

$$\mathbf{P} = \mathbf{X}(\Delta\mathbf{P}^0) + \mathbf{P}^0, \quad (3)$$

where \mathbf{P}^0 and $\Delta\mathbf{P}^0$ are the strings of the starting point parameters and their uncertainties, respectively. In this approach an individual \mathbf{G}_i can be identified with a genotype corresponding to all the simulation parameters $\mathbf{G}_i \rightarrow \mathbf{P}_i$. The associated phenotype is then constituted by the simulated EPR spectrum $\mathbf{Y}(B; \nu) = f(\mathbf{P})$, which is in turn compared with an environment — the experimental spectrum $\mathbf{y}(B; \nu)$ as it is shown in the box situated on the left-hand side in Fig. 1. The mean square (RMS) error was used as a fitness function to be minimized.

To launch the optimization process an initial population $\Omega(m) = \{\mathbf{G}_i\}$ of the individuals (potential solutions) was created at random from the assumed uncertainty range of the parameters. The population was next evolved following the GA principles, and better regions of the search space were explored by genetic operators of selection, crossover and mutation, that acted upon the initial population. A fitness proportional (roulette wheel) reproduction along with elitist succession was applied to produce a descendent sort of the individuals. A uniform crossover operator, consisting in arithmetic averaging of the parental chromosomes \mathbf{X}_1 and \mathbf{X}_2 was used for this purpose ($\mathbf{X}_1^{\text{new}} = \mathbf{X}_1^{\text{old}} + g(\mathbf{X}_2^{\text{old}} - \mathbf{X}_1^{\text{old}})$ and $\mathbf{X}_2^{\text{new}} = \mathbf{X}_2^{\text{old}} + g(\mathbf{X}_1^{\text{old}} - \mathbf{X}_2^{\text{old}})$, where g is a random number with uniform distribution in the $[0, 1]$ interval). A non-uniform mutation was described by the fluctuation function $\Delta(\tau, z) = z(1 - g^{(1-t/t_{\text{max}})^b})$, where t numbers the subsequent generations, z limits the maximum variation of the given gene to the specified

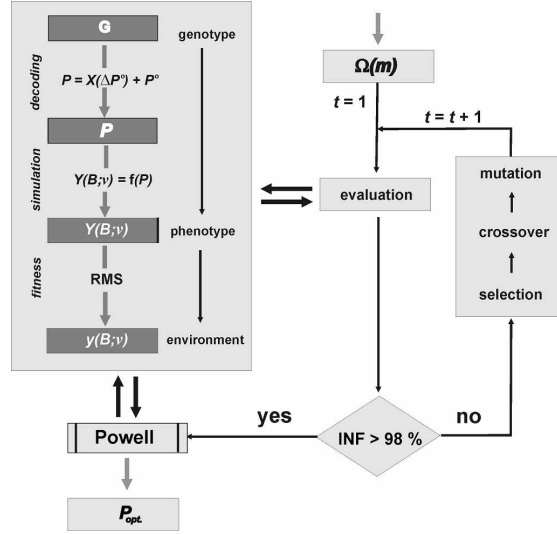


Fig. 1. Flowchart of the hybrid genetic algorithm used for optimization of powder EPR spectra. The relationship between the genetic space (genotype, phenotype, environment) and the problem space (parameters, simulated spectrum $Y(B; \nu)$, and experimental spectrum $y(B; \nu)$) is also shown.

range, whereas b describes the system non-uniformity [4]. During each generation t , the current population is improved in the sense that a fraction of inferior individuals is replaced by new ones with better fitness, obtained as the offspring (Fig. 1). The major strength of such algorithm is the ability to explore and exploit a large parameter space \mathcal{H}_p with no initial guesses and derivative information, and to avoid trapping in local minima. However, GA exhibits a distinct drawback, the true minimum, i.e., the best solution, is scarcely reached although the fittest individual in the final population should, in principle, be close to the optimal parameters [9]. Therefore, regardless of the quality of the results obtained with the genetic algorithm, deterministic Powell search was applied as a complementary measure to refine the ultimate results. An intimate neighborhood impact factor $INF(\mathbf{Y}, \mathbf{y}, \varepsilon)$ was used to trigger the swapping from GA to PA minimization, and to measure an overall quality of the fit (quantify the mutual resemblance of both spectra) as well [8]. The INF gauges the fraction of the simulated data points that are located within a user defined envelope surrounding the experimental spectrum of the widths ε (Fig. 2a).

3. Evaluation of the optimization algorithm

The efficiency of the hybrid implementation was evaluated using a complex test EPR spectrum with 17 adjustable parameters composed of three superimposed signals: an axial signal \mathbf{S}_1 with $g_{\parallel} = 2.37 > g_{\perp} = 2.06$ and $A_{\parallel} = 12.1$ mT,

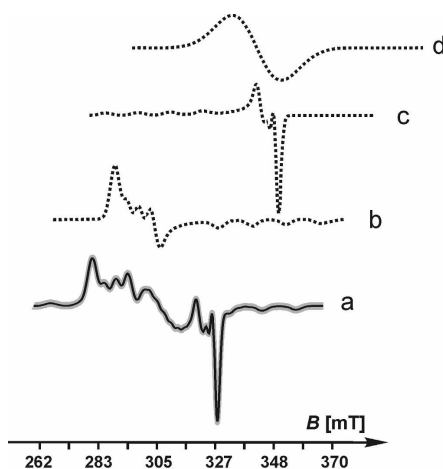


Fig. 2. Multi-component powder EPR test spectrum (a) along with its intimate neighborhood envelope, and (b–c) the constituent signals.

$A_{\perp} = 2.0$ mT, $\sigma_{\perp} = 2.3$ mT, $\sigma_{\parallel} = 4.5$ mT (Fig. 2b), a signal \mathbf{S}_2 with $g_{\perp} = 2.32 > g_{\parallel} = 1.97$ and $A_{\parallel} = 13.1$ mT, $A_{\perp} = 4.5$ mT, $\sigma_{\perp} = 4.0$ mT, $\sigma_{\parallel} = 4.0$ mT (Fig. 2c), and an isotropic signal \mathbf{S}_3 with $g_{av} = 2.18$ and $A_{av} = 16.0$ mT, $\sigma = 18$ mT (Fig. 2d). These parameters refer to various types of copper cage aquacomplexes in ZSM-5 zeolite [10]. Because all adjustable parameters are known *a priori*, in the case of successful optimizations a complete agreement between the calculated and fitted spectra can be obtained, identifying the global minimum univocally. For the convergence tests uncertainty limits $\Delta \mathbf{P}^0 = \pm 20\%$, $\pm 60\%$, $\pm 100\%$, and $\pm 120\%$ were imposed on all adjustable parameters, whereas the starting point \mathbf{P}^0 was selected at random. Eight independent initializations of the hybrid genetic algorithm were used to equilibrate the results. Following our earlier study on meta-optimization of the internal GA parameters [8], the crossover rate was set to $p_{crs} = 0.9$, the mutation rate $p_{mut} = 0.02$, and the population size $m = 200$ was used. As a fitting successfulness criterion we took $INF = 98\%$ for $\varepsilon = 0.028$. The results of the hybrid GA–PA search were next compared with the MC–PA optimization.

The performance profiles of the GA–PA search are shown in Fig. 3 for all the investigated ranges of $\Delta \mathbf{P}^0$. In these tests for successful optimization exact solutions ($RMS \rightarrow 0$) are in principle possible. From Fig. 3a–d it is apparent that all 8 runs of the genetic algorithm converged below the INF limit, regardless of the size of the assumed uncertainty range. The decrease in RMS was approximately exponential in nature, showing that the most rapid improvements occurred early in the evolutionary process. Although the overall progress was persisted as the algorithm evolves, its efficiency was remarkably slow down revealing an expected robustness of the global GA search and its inherent difficulty in fine-tuning. Therefore, once the convergence approached the INF level, the current values of all adjustable pa-

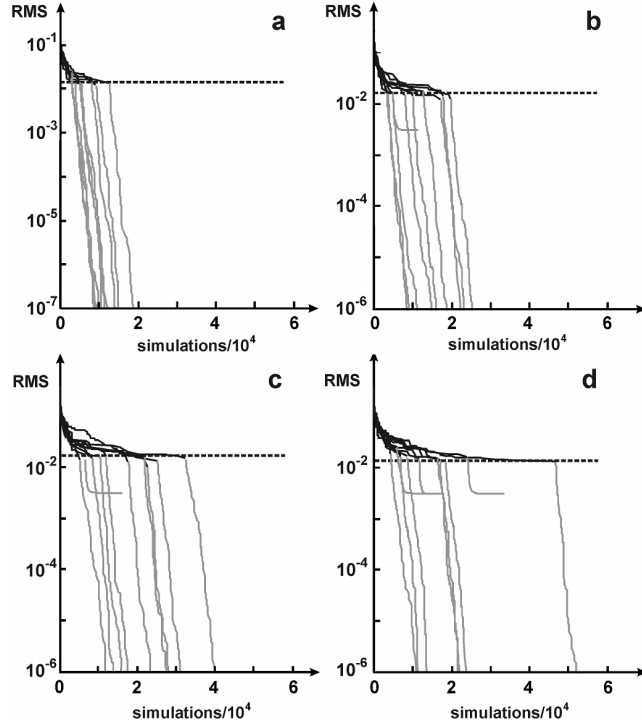


Fig. 3. Performance profiles of the hybrid genetic algorithm-Powell optimization for various range of input uncertainty (a) $\Delta P^0 = 20\%$, (b) $\Delta P^0 = 40\%$, (c) $\Delta P^0 = 60\%$, (d) $\Delta P^0 = 120\%$. The horizontal dashed line indicates an average RMS corresponding to $\text{INF} = 98\%$. Black lines indicate GA and grey — PA optimization steps.

parameters were taken as the starting point for the subsequent Powell refinement. Spectacular convergence, leading to a practically perfect fit ($\text{RMS} < 10^{-7}$) in most cases revealed that GA can very effectively locate promising regions in the spectral parameters search space, which are sufficiently close to the global minimum. The increasing uncertainty range ΔP^0 of the input parameters markedly influence the performance of the hybrid algorithm, as it can be inferred from the enhanced number of simulations required to successfully accomplish the fitting. It also somewhat deteriorates the subsequent PA refinements, which in one or two cases were caught in suboptimal regions, located yet well below the INF threshold (Fig. 3c, d). Thus, the ambiguity of the input parameters, influencing mainly the time of the minimization process, does not seem to be a very critical obstacle for achieving the final success.

The analogous progress of convergence using MC-PA optimization is shown in Fig. 4 for $\Delta P^0 = 40\%, \pm 60\%, \pm 100\%$, and $\pm 120\%$. The global search with MC method involved 2000 random trials of the algorithm for finding the most promising regions as the starting point of the subsequent PA fine tuning. To equilibrate

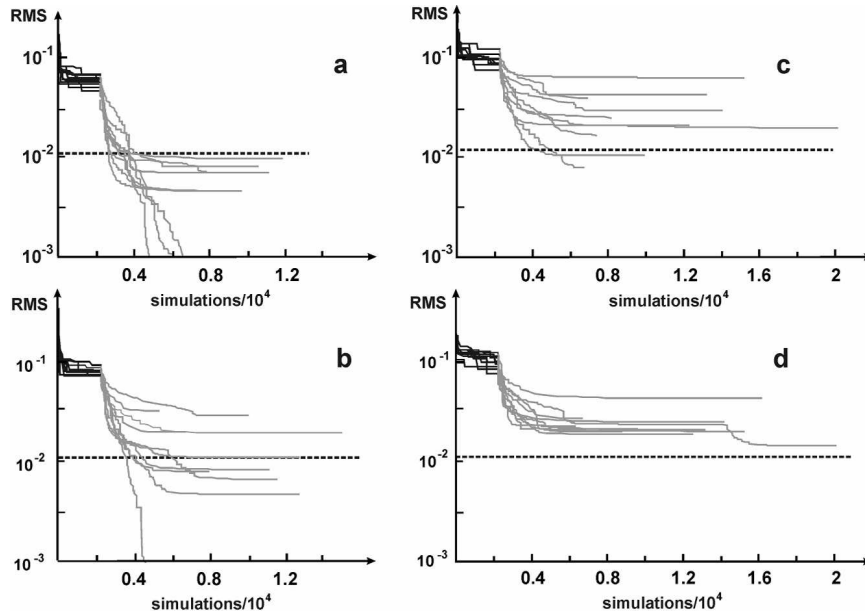


Fig. 4. Performance profiles of the Monte Carlo – Powell optimization for (a) $\Delta P^0 = 20\%$, (b) $\Delta P^0 = 40\%$, (c) $\Delta P^0 = 100\%$, and (d) $\Delta P^0 = 120\%$. The horizontal dashed line indicates an average RMS corresponding to $\text{INF} = 98\%$. Black lines indicate MC and grey — PA optimization steps.

the results ten independent MC–PA runs were executed. From the inspection of the results (Figs. 3 and 4) it is clear that the GA thoroughly outperforms the MC search. The global minimum was located only for the first two uncertainty ranges of the parameters, with 3 out of 10 successful trials for $\Delta P^0 = \pm 40\%$ (Fig. 4a), and only one successful trial for $\Delta P^0 = \pm 60\%$ (Fig. 4b). All other attempts became essentially stacked in various shallow minima located around the assumed INF limit, already at quite early stages of the optimization process. Further prolongation of the MC–PA optimization was completely unproductive in contrast to GA–PA search. For the large uncertainty interval of the parameters, $\Delta P^0 = \pm 100\%$ and $\pm 120\%$, each of the 10 runs failed to find the global minimum (Fig. 4c, d). All the trials leading to suboptimal solutions located, distinctly above the INF limit (with one exception) indicate that in such cases the MC–PA search was essentially unable to recognize the test spectrum properly. Thus, in contrast to GA–PA, for the MC–PA optimization subroutine the search efficiency was profoundly deteriorated while increasing the uncertainty range of the adjustable parameters.

4. Conclusions

The hybrid genetic algorithm combining global stochastic search with local deterministic refinement by the Powell method is capable of successful fitting com-

plex powder EPR spectra, even in the case of large uncertainties in the starting values of the adjustable parameters. The discussed examples revealed that owing to the joint action of the meta-optimized crossover and mutation operators, controlling the exploitation and exploration aspects of the optimization process, the genetic algorithm locates the most promising regions of the search space much more effectively than the conventional Monte Carlo method. The results demonstrate the promising power of the GA-PA approach in automated fitting and assigning of EPR spectra that can be hardly analyzed with conventional methods. The GA-PA method has been implemented into the simulation program EPRsim32 written in Microsoft visual C++ 6.0. The program runs under Microsoft Windows with full 32 coding and is available upon request.

Acknowledgment

This work was supported by the State Committee for Scientific Research, KBN, as the Research Project number 3 T09A 147 26.

References

- [1] J.R. Pilbrow, *Appl. Magn. Res.* **6**, 161 (1994).
- [2] F.E. Mabbs, D. Collison, *Electron Paramagnetic Resonance of d Transition Metal Compounds*, Elsevier, Amsterdam 1992.
- [3] R. Krzyminiewski, R.M. Kowalczyk, A. Bielewicz-Mordalska, Z. Pajak, P. Czarnecki, *J. Mol. Struct.* **471**, 243 (1998).
- [4] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Berlin 1992.
- [5] W.L. Meerts, M. Schmitt, G.C. Groenenboom, *Can. J. Chem.* **82**, 804 (2004).
- [6] J.R. Pilbrow, *Transition Ion Electron Paramagnetic Resonance*, Clarendon Press, Oxford 1990.
- [7] A. Rockenbauer, *Mol. Phys. Rep.* **26**, 117 (1999).
- [8] T. Spalek, P. Pietrzyk, Z. Sojka, *J. Chem. Inf. Model.* **45**, 18 (2005).
- [9] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, New York 1989.
- [10] B. Gil, J. Datka, S. Witkowski, Z. Sojka, E. Broclawik, *Stud. Surf. Sci. Catal. D* **130**, 3249 (2000).