GENETIC ALGORITHM APPLIED TO MULTILAYER STRUCTURE DETERMINATION

Daniel Šimek[1], David Rafaja[1], Jiří Kub[2]

[1]Faculty of Mathematics and Physics, Charles University in Prague
E-mail: simek@apollo.karlov.mff.cuni.cz[a], rafaja@karlov.mff.cuni.cz
[2]Institute of Physics, Academy of Sciences of the Czech Republic
E-mail: kub@fzu.cz

Abstract

This paper shows an alternative way of the multilayer structure refinement to the well established least-squares method. If calculated for the low-angle x-ray reflectivity, the residual function (based on squared differences between the measured and the calculated reflectivity curves) is oscillating with respect to some parameters. This makes the refinement by the linearised least-squares method complicated and ineffective. Genetic algorithm can be an efficient substitute or at least a complementary technique to the linearised least-squares method.

Keywords:

Genetic algorithm, multilayers, X-ray reflectivity

1. Introduction

X-ray scattering from multilayers involves the specularly reflected wave and the diffuse scattering [1]. The first author, who observed the dynamical effects in the diffuse scattering was Y. Yoneda [2]. Intensity of the specularly reflected (coherent) wave approaches the incident wave intensity, therefore the coherent wave has to be treated dynamically. The technique used to measure the reflectivity is a symmetrical $\theta/2\theta$ scan; the typical 2θ angles are of order of few degrees. In the specular reflectivity, so called Bragg-like maxima and Kiessig oscillations can be observed (see Fig. 1). The Kiessig oscillations appear if the whole multilayer stack has a well-defined thickness and (relatively to its thickness) a low roughness at the surface and at the substrate. The Bragg-like maxima (at 1.1, 2.0 and 3.0° in Figure 1) arise in case of periodic multilayers; their angular positions can be calculated from the Bragg law, which must be modified to respect the refractive index of the matter. Instead of the interplanar distance, the thickness of the repeated motif stands out in the modified Bragg equation.

Apart from the coherent wave, the diffuse scattering is also produced, since there are many imperfections in the multilayer microstructure. The diffusely scattered wave is at least few orders of magnitude weaker; thus the kinematical approach is quite justified. Unlike for the kinematical theory of diffraction, there is not only the strong incident wave present in the low-angle scattering, but also the specularly reflected wave has to be considered. This leads to the well-established semi-kinematical approach referred to as DWBA (Distorted Wave Born Approximation) [3].

The imperfections can be generally divided into the disturbances at the atomic scale, to which the low-angle x-ray scattering is not sensitive at all, and into the well-correlated microstructure imperfections. The former just decreases the specular reflection, the latter too, of course, but it is also responsible for the diffuse scattering.

The scans used to measure the diffuse scattering are mainly the ω-scans and the so-called off-scans (this means $\theta/2\theta$ scans with a non-zero ω offset from the symmetric position). Another type of the asymmetrical scan is the detector scan (2θ-scan), but this was not used in our case. In the specular scan (symmetrical $\theta/2\theta$ scan), there is not only the specular reflectivity (coherent wave) present, but there is also a portion of the diffuse scattering, which leans right into the detector slit direction. We can reduce this portion by approving the resolution (by narrowing the primary beam divergence and detector angular resolution), but it can never reach zero. The specular wave has a finite angular divergence due to the sample curvature and block-shaped multilayer structure; thus the intensity ratio between the coherent wave and the diffuse scattering is finite in any direction. There is a possibility to neglect the diffuse scattering in the specular scan for very small angles. However, as the reflection angles increase, the specular wave diminishes and the diffuse scattering in the specular direction must come into account.

The microstructure parameters of the multilayers, which can be investigated by measuring the reflectivity and low-angle diffuse scattering, are as follows:

- Thickness of the whole multilayer system and thickness of the repeated motif (in case of a periodic multilayer)
- In some cases also the thickness of several layers (if there are not many)
- Mean electronic density of the layers (relatively to the reference structure)
- Geometrical roughness and interfacial interdifussion depth rms values
- Correlation parameters of the geometrical roughness (lateral correlation length and the degree of vertical replication)

2.   Motivation

Reflectivity curves contain most of the structural information on the multilayer. Those are the period (if we suggest a periodic multilayer), total roughness of the interfaces (this includes both the geometric roughness and the interfacial interdiffusion) and the mean electronic densities in the layers [4].

If one regards only the case in which all nominally same layers have the same thickness and electronic density, then the fit of the calculated curves is not satisfactory. Moreover, a precious information on the technology imperfection is lost. The deposition conditions often change during the deposition process, and a trend or simply deviations can then be obtained from the multilayer structure parameters.

If we assume that all parameters are independent, we have to treat another problem: many partially correlating parameters. Saying "partially correlating" we mean that the residual function minima are enlarged in a general direction. A minimum with respect to one parameter then shifts with a change of another parameter, i.e. when looking for a minimum, there is a need to change more parameters at once, not just one after another. Let us suggest a change of the thickness of a single layer by the value of about the period (period is the thickness of the repeated motif). Then the main Bragg-like maxima do not move at all (Fig. 1). If we replace the corresponding increase of the total multilayer thickness by decreasing (formally) the thickness of the next two layers, the calculated reflectivity curve is almost the same. Not exactly, sure, but we have modified one of the top-most layers, to which the reflectivity is most sensitive. The difference can be observed, but if both the original and the theoretic reflectivity curves of the modified model were compared with the measured curve, no one could know, which one would be better.
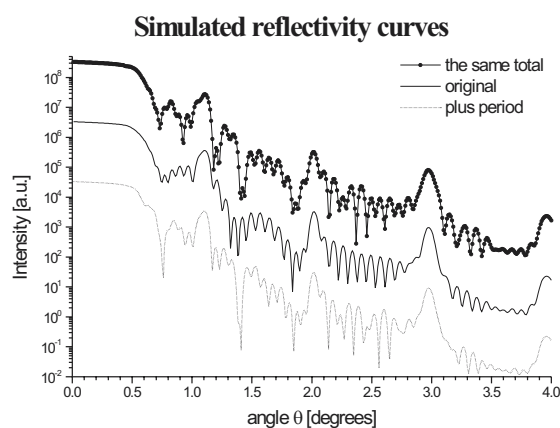

Figure 1. Simulated reflectivity curves for different thickness of a single layer. Original structure is [20 Å Fe / 25Å Au]×10 on Si with 100 Å Au cap (solid line). Then: second Fe layer from top modified to 65 Å (dotted line), and finally all neighbouring Au layers thinned to 2.5 Å (solid line with circles in calculated points). One curve is divided by 100 and the other is multiplied by 100 for better recognition.

Thus, applying the non-linear least-squares method on the initial not-very-perfect model, the structural parameters can leave the rational limits. Consequently, any other step of the algorithm needn't find better model describing the real structure more reliably.

This is caused by the oscillatory behaviour of the residual function. The least squares method simply slides straight into the nearest minimum of the residual function and won't leave it any more.

3.   Solution of the problem

The problem described in the previous part can be treated by means of the genetic algorithm, especially when this is combined with the least-squares method.

Usually, if the least-squares method fails as described above, it can be mended by starting the refining process with a little different initial parameters. This may produce better but also a worse solution. If one has a lot of patience, finally the best solution can be obtained.

Another approach is to use the genetic algorithm for random creation of the initial parameter sets. It could be used just to produce random parameters spread around the original model, which are refined over and over again, storing every better solution than the previous best one can yield. This technique does work but it is quite exaggerating to call it genetic algorithm.

The real genetic algorithm's power insists not only in producing the random initial parameters for the least-squares method, but also in capability to refine them.

Indeed the term "refine" is not accurate, when we are talking about the genetic algorithm. Genetic algorithm cannot improve the parameters by modifying them in the direction of better solution like the derivative-methods do. However, we are going to use the term "refinement" in a figurative sense to denominate its capability to find better solutions by applying itself repeatedly.

4.   Mathematical background

Upon the least-squares method one minimises the weighted sum of squared differences between calculated and measured intensities:

$$f = \sum_i w_i f_i = \sum_i w_i (y_i^{calc} - y_i^{meas})^2, \qquad (1)$$

where i stands for experimental points. In case of reflectivity curves, the intensity (y) varies by several orders. The weight w_i has to be adjusted to the fact:

$$w_i = (1/ y_i^{meas})^2, \qquad (2)$$

which corresponds to logarithmic scale refinement with a constant weight (for small differences):

$$f = \sum_i f'_i = \sum_i \left[\ln(y_i^{calc} / y_i^{meas})\right]^2. \qquad (3)$$

In order to make the genetic algorithm co-operating with the least-squares method (note we are attempting to combine both methods, not just to apply them one after the other), it has to look for the best solution by the same rules, i.e., genetic algorithm has to minimise the same function (3) by changing the free parameters.

In terms of the genetic algorithm [5], one parameter is equivalent to a *gene*. Genetic algorithm works not only with one set of the parameters (the *individual*), but it stores several different sets – the *population*. The population is developed by means of *genetic operators*.

If we represent the individual by a vector a, its classification (also referred to as *fitness function*) $f(a)$ is a scalar, then the whole population forms the matrix A and a classification vector $f(A)$. We are actually minimising the classification function (3) instead of maximising the fitness capability of the individuals.

There are two main ways of developing the population – *mutation* and *crossing-over*. The mutation is a random process, which is basically responsible for a population gene range spreading. A simple mutation insists in a random change of a randomly selected parameter a_j, which belongs to a randomly chosen individual a. The change, in our case, happens around the initial value (so called boundary mutation [5]) with normal (Gaussian) distribution with a width defined as a function of the initial parameters:

$$b_j = a_j + \text{rnd}(\text{mute}_j(a));$$
$$b_k = a_k \quad k \neq j$$

(4)

Function *mute*$_j$ determines the mutation range depending on the starting values of the free parameters and may be considered as a constant for instance; $rnd(x)$ represents a random number with normal distribution around zero:

$$w(\text{rnd}(x)) = \exp\left(-\frac{\text{rnd}(x)}{2x^2}\right).$$
(5)

The new created individual b is either added to the population or it replaces any other individual (e.g. the worst one), that is *killed*. The only need to keep the genetic algorithm working is to preserve the best individual from killing.

The cross-over operator works upon two different randomly or even systematically selected individuals a and b. The new individual is a linear combination of those original ones with a weight inversely proportional to their classifications:

$$c_0 = \frac{a \cdot f(b) + b \cdot f(a)}{f(a) + f(b)}.$$
(6)

If the initial individuals lie in opposite direction with respect to the global minimum, this process can approach the minimum, otherwise not. That is why it is useful to combine this process with a random mutation in all parameters around this "centre of gravity" with a normal distribution widths equal to parameter differences of the original individuals:

$$c = c_0 + \text{rnd}(a - b),$$
(7)

where $rnd(x)$ is a vector of rnd(x). Thereby, the parameters can also leave the initial range in the better individual's direction.

In the matrix formalism presented above, each genetic operator represents a matrix multiplication of a population matrix A by the evolution matrix D and an addition of the mutation matrix M. Let us suggest that the matrix A contains the individuals in rows. Then:

$$A^{new} = D \cdot A^{old} + M.$$
(8)

Matrix D is of the size $n \times n$ if the new individual replaces another or $n \times (n+1)$ if the new one is added to a former n-member population. In our case (only one new individual is produced), the evolution matrix is almost the identity one, except a single row (index j, where j is the individual to be killed or $j = n + 1$). Anyway, all rows have sum of the components equal to unity. The D-matrix of the *mutation* operator is equal to the unity matrix. The *cross-over* operator has two non-zero components in j-th row: $D_{jk} = f_l/(f_k+f_l)$ and $D_{jl} = f_k/(f_k+f_l)$ (k and l are different random indices 1 to n). Let us remind that $f(A)$ is a classification column vector. The other rows are the same as in the unity matrix. Mutation matrix M has only one non-zero row, namely the j-th one. Its values have been discussed above.

Now, any application of genetic operators produces so-called new *generation* with just a single different individual. However, there is still a need to calculate the new individual's classification, i.e. the j-th component of the classification vector f. This means the calculation of the simulated scattered intensity in all experimental points and then the differences according to Eq. (3). Since most of the computational time is consumed by the evaluation of the new individuals' classification, there is no progress in producing more than one new individual in every generation, i.e. more complicated expression for evolution and mutation matrices D and M.

5.   New individuals generation combined with the least-squares method

Since the genetic algorithm cannot treat the partially correlating parameters, e.g. the root mean square roughness of several interfaces, this method cannot be satisfactory itself. A change of a parameter (even the right way) may produce worse solution than the original one. Genetic algorithm is unable to recognise any correlation of parameters and that is why the combination with the least-squares method is useful.

The process insists in refining every new created individual immediately after it arises. After a certain number of refining cycles the new individual is classified for the last time and the next generation is completed. Then the new individual should lie in a local minimum of the residual function, at least with respect to the parameters refined by the least squares method.

The choice of parameters to be refined by the one or the other (or even by both) methods depends on their definition. In case of our program, the real multilayer structure

parameters are arbitrary "user-defined" functions of the refined parameters. Refined parameters are then just variables without physical meaning. Using this approach, some kind of correlation can be reduced. For example, if one defines a mean period (thickness of repeated motive), period perturbations (for several layers) and a fractional part of every layer thickness in the repeated motif instead of defining several layers' thickness as independent parameters, the correlation is partially reduced.

In our implementation, the population has the maximum number of individuals given by the user. Coming up from the initial model, the new individuals are added until the number reaches the limit. Then, the new created individual always replaces the worst one. This has proven not to be the best choice, because when all the individuals fall into one local minimum, the only way to reach the next one is through a single mutation. That is to say that any individual with values of the parameters between these minima is classified much worse and the new individual is killed in the next step. There is still a chance for it to mutate right in the next step, but the probability is very low.

The function "mute" (mentioned in chapter) is set by the user in accordance with the definition of refined parameters. It has to enable reaching the nearest minima in classification function through the one step, as we have discussed before.

There is also a possibility to choose, which scans (experimental data) are used for evaluation of the classification function. Such an approach is necessary if the refined set of parameters does not involve any, to which some scans are highly sensitive. These scans then have to be excluded from the refinement, because they could undesirably influence the values of refined parameters. This effect comes up from the fact that the omitted scans are usually less sensitive to some of the refined parameters. If they were used, they would force them to change the wrong way in order to compensate wrong values of the non-refined parameters.

6. Refining process

Now we start to present the refining process on the structure parameters of two different specimens. The first one is (nominally) the molybdenum monolayer on Si substrate.

The substrate was first polished by ion beam. The cleaning process was responsible for deposing a thin tungsten layer, which originated from the tungsten cathode.

We tried this simple example to prove the ability of the genetic algorithm to find a reliable model of the structure, even if the first model is completely wrong. As we wrote before, one of the parameters most complicated to be refined by the least-squares method is the layer thickness. It is connected with oscillations in the reflectivity curve, and their period is inversely proportional to the thickness of the layer. If we initially set the thickness to be much less than it really is, the theoretic curve is at the whole different (Fig. 2). A slight change of it even the right way will produce an inestimable change of the residual function. Thus, the least-squares method is unusable for the case.

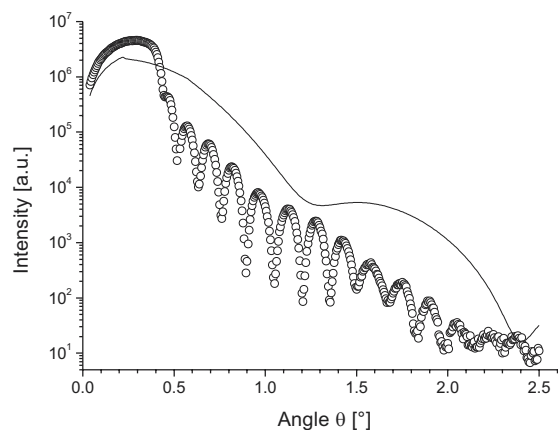Using a genetic algorithm with an adequately large mutation range for thickness, the new created individuals

Figure 2. Reflectivity curve of a molybdenum and tungsten bilayer on silicon (open circles). Theoretic curve (solid line) is calculated for $d_{Mo} = 20$ Å and $d_W = 18$ Å.

spread their thickness in a large scale of values, as the population grows. The limit number of individuals was set to 100. Then the crossing-over process and also the mutations vary the thickness until the good agreement with the real curve is gained. Such an individual is necessarily the best one and won't leave the population until there is some better one.

Now it pays to change the mutation range to quite a smaller value, not exceeding the period. This can be also satisfied by setting the thickness' "mute" function to be proportional to the thickness and by starting the refinement with the large value of thickness. After a certain amount of time, the individuals with very different thickness dispose, as they are replaced by the new created better ones, which are closer to the right model.

Finally, there are many individuals with only a little different parameters. The best one is considered to be the "best fit" model of the structure (Figure 3); its parameters are shown in Table 1. Note that the real thickness of the layer is one order different from the initial model.
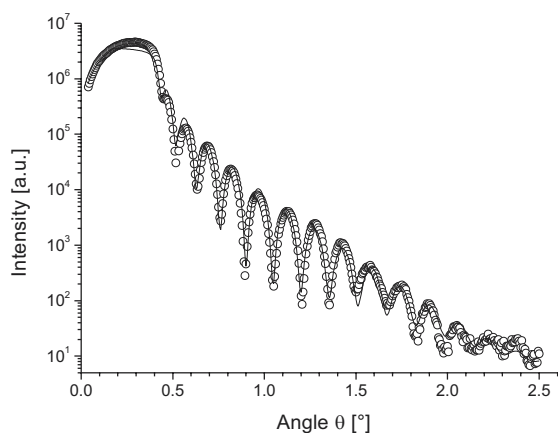

Figure 3. Reflectivity curve of a molybdenum and tungsten bilayer on silicon (open circles) and refined theoretic reflectivity curve (solid line).

Table 1. Molybdenum and tungsten bilayer refined parameters

	ρ	d [Å]	σ_{eff} [Å]
Mo	0.88	265	8.6
W	0.81	7.0	9.0
Si	1.00		6.5

Together with the thickness, all the other parameters were refined. Those were the relative electronic density (with respect to the bulk material) and effective interface roughness (of the upper interface). There were also some non-structural parameters refined, such as sample length, curvature, primary beam intensity and background.

The second specimen is a multilayer with the following nominal parameters: On the silicon substrate a multilayer system is deposited. It consists of 8 Fe/Gd bilayers (Fe is first at the substrate). On the top there is an Al cap layer.

Thickness of several layers was first examined by means of High Resolution Electron Microscopy (HREM), thus the first approach was rather good. The initial shape of the theoretical curves is compared with that of the experimental ones in Figure 4.
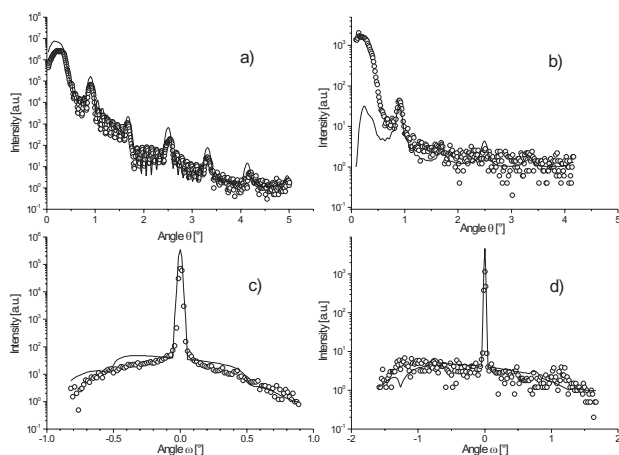

Figure 4. Measured curves (open circles) in comparison with the curves calculated accordingly to the initial model (solid lines). a) specular scan (reflectivity), b) off-scan with the offset from specular position $\omega = 300"$, c,d) omega-scans with $\theta = 3240"$ and $6000"$.

In the first step, only the mean bilayer thickness (the period), mean thickness ratio (of Fe and Gd layer) in each bilayer and the thickness of the cap layer were refined together with the roughness of the lowest interface and its mean interlayer increase (every upper interface was supposed to have a greater rms roughness). Also some non-structural parameters, such as sample length, curvature, etc., were refined. The only scan used for evaluation of the classification was the reflectivity curve, because of its sensitivity to these parameters (Figure 5).

Then the geometrical part of the roughness (caused by the real rough surface, not by the imperfections at the
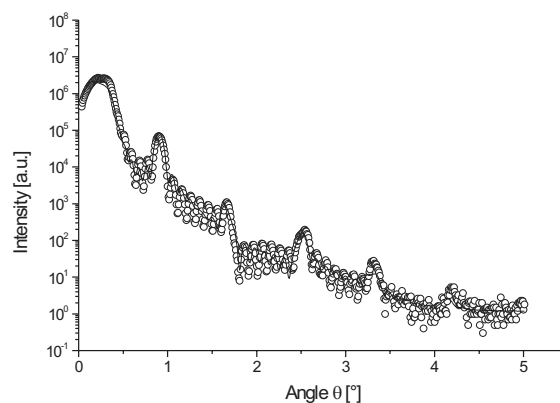

Figure 5. Simulation of reflectivity curve after first refinement of main parameters.

atomic scale) and its correlation parameters were refined together with the layer thickness. The density aberrations and the roughness of several interfaces were treated independently. That time, all types of scans had to be used in classification, because only the diffuse scattering distinguishes the kind of imperfections and is sensitive to their correlation parameters. The final fit is shown in Figure 6. The final multilayer parameters were obtained as shown in Table 2. For the multilayer system, the mean values are shown together with their mean statistic deviations. The means are calculated over the whole multilayer stack. Relative density of the substrate was considered to be equal unity as a reference value, which needn't be too accurate, as is obvious, if we take a look at the Fe and Gd relative density mean values.

Geometrical roughness decreases from the bottom to the top, starting at about 5 Å at the substrate and ending with values comparable with the mean thickness deviation (0.4 Å). Geometrical roughness then increases again through the last 5 layers up to about 10 Å at the surface. The latter effect corresponds probably to the surface oxidation.
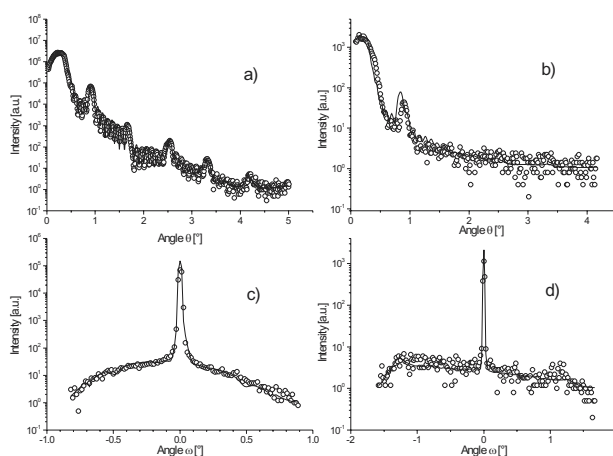

Figure 6. Measured curves (open circles) in comparison with the final theoretic curves with the maximum of free parameters (solid lines) – the "best fit". a/b/c/d – see Figure5 caption.

Table 2. Refined values of the main multilayer parameters. Deviation is the rms difference from the mean value for several layers of the same type.

	ρ		d [Å]	
	value	deviation	value	deviation
Al (oxidised)	0.61		19.4	
Al	0.86		13.1	
Gd	1.07	0.02	27.5	0.4
Fe	1.16	0.02	25.7	0.4

The other part of the effective roughness (so called the diffuse one) is nearly the same throughout the system, greater at the interfaces with Gd below (about 6 Å) and less at the interfaces with the Gd above (about 3 Å). It is the dominant component of the roughness, the evidence of which is given by the low level of diffuse scattering.

The mean vertical correlation coefficient in Fe layers is about 0.82, in Gd layers about 0.76. Greater values for iron layers were expected due to that the Fe layers are thinner than the Gd ones. The correlation falls down near the surface. Lateral correlation length is about 300 Å for all interfaces.

7.　Conclusions

Genetic algorithm is a powerful method in the refinement of the multilayer structure parameters. Its main convenience is, in comparison with the least-squares method, in the stability of calculation. Even in combination with the least-squares method, when this fails because of the singular normal matrix, or if the refinement is not convergent, the genetic algorithm can treat the new individual without any further refinement. The process can then continue and no user interference is necessary.

In comparison with our program MUSIX recently used for the reflectivity curves calculation, which uses only the least-squares method, the probability that the structure parameters leave the rational limits is much lower for the genetic algorithm. This is brought on by the fact that the best individuals are always left unchanged when producing the new generation. Any non-rational structure in general has to be worse then some reasonable in terms of the classification function and thus may even be disposed from population later.

The computational time is, alas, longer, but user's time is saved. The more time it is given for calculation, the better the solution. The new approach insists mainly in the joint refinement of the reflectivity curves together with the low-angle diffuse scattering curves, collected in different types of scans. There is only one set of parameters used for the calculation of all the curves and only the joint refinement is the pure technique to treat the effects, which cannot be separated by experimental means.

References

[1] M. Jergel, V. Holý, E. Majková: *Thin Film Studies by the X-ray Reflectivity and Diffuse Scattering at Grazing Incidence*, Material Structure 3 (1), 101 (1996).

[2] Y. Yoneda: *Anomalous Surface Reflection of X-rays*, Phys. Rev. 131, 2 010 (1963).

[3] S. K. Sinha, E. B. Sirota, S. Garoff, H. B. Stanley: *X-ray and Neutron Scattering from Rough Surfaces*, Phys. Rev. B 38, 2 297 (1988).

[4] V. Holý, T. Baumbach: *Nonspecular X-Ray reflection from rough multilayers*, Phys. Rev. B 49, 10 668 (1994).

[5] D. E. Goldberg: *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, Massachusetts (1989).

[6] Xiaochun Hu: *A Survey of GA Solutions for Optimization with Constraints*, presently at http://www2.ics.hawaii.edu/~sugihara/course/ics691f98/xiaochun-survey.htm