II

Publication II

J. Tiilikainen, V. Bosund, J.–M. Tilli, J. Sormunen, M. Mattila, T. Hakkarainen and H. Lipsanen, *Genetic algorithm using independent component analysis in x–ray reflectivity curve fitting of periodic layer structures*, Journal of Physics D: Applied Physics **40** (2007) 6000–6004

IOP Publishing Journal of Physics D: Applied Physics
J. Phys. D: Appl. Phys. **40** (2007) 6000–6004 doi:10.1088/0022-3727/40/19/033

Genetic algorithm using independent component analysis in x-ray reflectivity curve fitting of periodic layer structures

J Tiilikainen, V Bosund, J-M Tilli, J Sormunen, M Mattila, T Hakkarainen and H Lipsanen

Micro and Nanosciences Laboratory, Helsinki University of Technology, Micronova, PO Box 3500, FI–02015 TKK, Finland

E-mail: jouni.tiilikainen@tkk.fi

Received 25 June 2007, in final form 6 August 2007
Published 21 September 2007
Online at stacks.iop.org/JPhysD/40/6000

Abstract
A novel genetic algorithm (GA) utilizing independent component analysis (ICA) was developed for x-ray reflectivity (XRR) curve fitting. EFICA was used to reduce mutual information, or interparameter dependences, during the combinatorial phase. The performance of the new algorithm was studied by fitting trial XRR curves to target curves which were computed using realistic multilayer models. The median convergence properties of conventional GA, GA using principal component analysis and the novel GA were compared. GA using ICA was found to outperform the other methods with problems having 41 parameters or more to be fitted without additional XRR curve calculations. The computational complexity of the conventional methods was linear but the novel method had a quadratic computational complexity due to the applied ICA method which sets a practical limit for the dimensionality of the problem to be solved. However, the novel algorithm had the best capability to extend the fitting analysis based on Parratt's formalism to multiperiodic layer structures.

1. Introduction

X-ray reflectivity (XRR) is a fast tool for noncontact thin film metrology, which provides information on properties such as film thickness, mass density and interface/surface roughnesses. The measurement produces an oscillating curve which can be directly analysed by Fourier transform [1] or wavelet analysis [2, 3]. These methods are based on kinematical approximation and are not as accurate as analyses based on Parratt's formalism [4] combined with Nevot–Croce interface roughness approximation [5]. Parratt's formalism, however, requires curve fitting. The fitting cannot be done efficiently only with gradient, simplex or simulated annealing methods due to multiple local optima in the search space or huge problem dimensionality. This fitting can be efficiently performed by genetic algorithms (GAs) [6–10] but due to the computationally divergent nature of the applied formalism, some additional techniques are needed to obtain robust convergence. We showed in our previous paper that interparameter dependences, called genetic linkage, decrease the fitting performance of a simple algorithm in XRR curve fitting and these can be reduced by the rotation of coordinates [10] in the mating phase. The earlier success of the covariance based statistical approach in this phase suggests that more sophisticated statistical techniques, such as an independent component analysis (ICA) can enhance converge properties further. In [11, 12] the success of evolutionary algorithms using ICA in linkage reduction was addressed with a few synthetic problems but this approach has not gained more attention.

In this paper we study which linear transformation technique used in the mating phase offers the best improvement in the median convergence of a GA. The median convergence of the novel independent component analysis GA (ICAGA) is compared with principal component analysis GA (PCAGA) and to conventional GA (CGA). The algorithms are tested by fitting trials to a theoretical XRR curve based on the realistic model of a periodic multilayer structure. It is shown that CGA is the best algorithm in the simplest fitting cases while

PCAGA and ICAGA have nearly equal performance with moderately sized problems. ICAGA is shown to have the best median convergence properties with the most difficult XRR fitting problems. In section 2, the importance of the genetic linkage reduction is discussed and the principles of PCA and ICA techniques are presented. In section 3, the test problem is introduced. The implementations of GAs are presented in section 4. Finally, the fitting performance of CGA, PCAGA and ICAGA are tested as the function of problem dimensionality in section 5.

2. The reduction of interparameter dependences

Interparameter dependences, called genetic linkage in combinatorial optimization, means that the parameters have no clearly independent contribution to the fitness to be minimized but the contribution of one parameter depends on the others. Dependences are problematic in a conventional GA since parameters are optimized separately resulting in inefficient optimization. This can be circumvented somewhat if linear transformation is used to modify coordinates so that interparameter dependences are reduced. This procedure is also called the separation or the demixing of sources and it is applied here for XRR fitting parameters.

The demixing (or separation) matrix $\boldsymbol{M}^{-1}$ used in the linear transformation is calculated from a sampling set which consists of selected trials computed earlier. Let us denote the ith trial describing the information of $n \geqslant i$ layers by

$$\boldsymbol{x}_i = [\boldsymbol{t}_i, \boldsymbol{m}_i, \boldsymbol{r}_i], \tag{1}$$

where $\boldsymbol{t}_i$ is a $1 \times n$ row vector representing the thicknesses of layers, $\boldsymbol{r}_i$ is a $1 \times n$ row vector representing the roughnesses of layers and $\boldsymbol{m}_i$ is a $1 \times n$ row vector representing the mass densities of layers. Then $m \times 3n$ sampling set matrix is defined as

$$\boldsymbol{X} = \left[\boldsymbol{x}_1^{\mathrm{T}} \boldsymbol{x}_2^{\mathrm{T}} \ldots \boldsymbol{x}_m^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{2}$$

where m is the number of samples. The demixing matrix can be computed inexpensively by PCA which removes the correlation between the columns, i.e. parameters. Then mixing matrix

$$\boldsymbol{M} = \mathrm{Eig}[\mathrm{Cov}(\boldsymbol{X})], \tag{3}$$

where Cov ($\boldsymbol{X}$) means the computation of covariances between the columns and Eig calculates normalized eigenvectors, which are the columns of $\boldsymbol{M}$. Using the transformation

$$\boldsymbol{y}_i = (\boldsymbol{M}^{-1} \boldsymbol{x}_i^{\mathrm{T}})^{\mathrm{T}}, \tag{4}$$

each component of $\boldsymbol{y}_i$ belongs to a distribution which is not correlating with other distributions. This and other transformed trials are used in combination process in GA.

One of the main limitations of PCA is related to the orthogonality of the basis which is not necessarily required in ICA. Whereas the separation matrix in PCA is determined by the eigenvalue decomposition of a covariance matrix, the demixing matrix in ICA is determined by minimizing mutual information between variables using an appropriate linear transformation. Intuitively speaking, the minimization of mutual information corresponds to the separation of original distributions to parameters distributed in as nongaussian a way as possible [14]. The justification of this approach relies on the central limit theorem, which says that the sum of any set of distributions is more Gaussian than the originals. The computation of a demixing matrix using this scheme is not, however, straightforward. There are several definitions of the measure of mutual information and several iterative techniques to determine the separation matrix which naturally means different computational efficiency and robustness between different ICA methods.

In this study we use the efficient FastICA (EFICA) method developed by Koldovský *et al* which was shown to be fast and robust [16]. EFICA is applied to determine the separation matrix iteratively using a previously determined demixing matrix as the starting point. In the first cycle of ICAGA, there is no previously determined separation matrix available so the initialization is based on random numbers. After an iteration step, the separation and mixing matrices obtained from EFICA are modified so that the length of the projecting vectors are normalized to preserve the space resolution between the original and the transformed spaces. The sampling set for PCA and ICA is selected to be all the previously calculated individuals to reduce finite sampling problems.

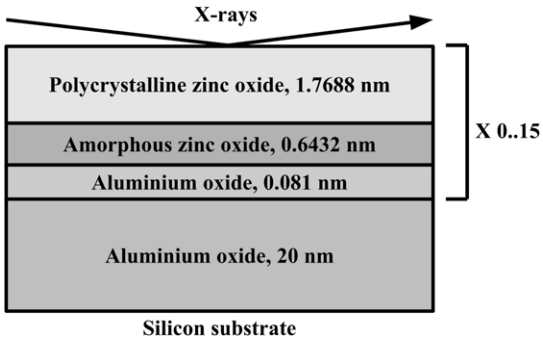

Figure 1. Schematic structure of the modelled nanolaminate with varying number of thin ZnO/AlO-periods.

3. The problem setting

The performance of three GAs was tested with XRR curve fitting tests. Target XRR curves were computed based on models corresponding to realistic atomic-layer-deposited (ALD, [13]) structures. Figure 1 shows the case considered in this paper. This theoretical structure, based on the results shown in [15], was utilized to examine the performance of the algorithms with an increasing number of periods. Table 1 summarizes the parameters used in the layer structure.

The presented layer structure was utilized to calculate XRR curves based on Parratt's formalism and Nevot–Croce roughness approximation. The XRR curve angle range was 0°–2.5° and 200 evenly spaced points were calculated for the target curve. The same number of data points were also calculated for trial curves but these curves were based on the physical parameters which were randomly deviated ±20% from the target parameters. Fixed interparameter dependences were not used. The parameters were afterwards normalized between [0, 1] so that 0.5 corresponds to the target value. The dimensionality of the problem was $2+3+n \times (3 \times 3) = 5+9n$

Table 1. The parameters of the layer structure used in simulation. The plus sign with roughness means that interfacial roughness is cumulatively increased with every layer on silicon substrate. Negative sign means that interfacial roughness is decreased due to an etching effect.

Material	Mass density (g cm^{-3})	Roughness (nm)
Polycrystalline ZnO	5.52	+0.07
Amorphous ZnO	4.32	+0.01
Thin Al_2O_3	3.0	-5×10^{-5}
Al_2O_3	3.0	+0.5
Silicon	2.33	0.3

(the mass density and roughness of the substrate and all the parameters of Al_2O_3 and n periods of the repeating structure).

4. The implementation of the algorithms

4.1. Implementation of the genetic algorithms

The implementation of the novel ICAGA for XRR curve fitting was based on CGA and PCAGA. All the features of ICAGA and PCAGA utilized the same code and the same internal parameters except the determination of a new mixing matrix and the formation of the subsequent population. The structure and details of the implementation is also presented in our previous paper [10], where PCAGA is denoted are CovGA without nonlinear-fitness-space-structure adaptation and some other functionalities.

Figure 2 shows the implementation used in the GAs. The initial population was generated using the uniform random distribution. The fitness function F defined as

$$F = \left\{ \sum_{i=i_c}^{N} \left[(x_{i,t} - x_{i,g})^2 x_{i,g}^{-2} \right] \right\}^{1/2} \tag{5}$$

was applied, where $x_{i,t}$ and $x_{i,g}$ denote logarithms of one of the N datapoints of the trial and the target curves, respectively. Index i_c corresponds to the critical angle of the target curve. Each cycle in GA includes the following steps:

(i) The best individual in the population is selected to mate with k other randomly selected individuals.

(ii) PCA or ICA is performed to compute mixing and demixing matrices. The parents are mapped using the separation matrix

$$\boldsymbol{y}_{\text{parents}} = (\boldsymbol{M}^{-1}\boldsymbol{X}^{\mathbf{T}}_{\text{parents}})^{\mathbf{T}}. \tag{6}$$

(iii) A crossover operator is used in the creation of new individuals

$$\boldsymbol{y}_{\text{child},i} = \boldsymbol{C}\boldsymbol{y}_{\text{parent,elitist}} + \text{Not}(\boldsymbol{C})\boldsymbol{y}_{\text{parent},i}, \tag{7}$$

where the uniform crossover operator C selects random genes from the elitist parent and Not(C) the rest from $\boldsymbol{y}_{\text{parent},i}$ in proportion to the parents' fitnesses. The offspring are

$$\boldsymbol{y}_{\text{offspring}} = [\boldsymbol{y}_1^{\mathbf{T}}\boldsymbol{y}_2^{\mathbf{T}} \ldots \boldsymbol{y}_k^{\mathbf{T}}]^{\mathbf{T}}. \tag{8}$$

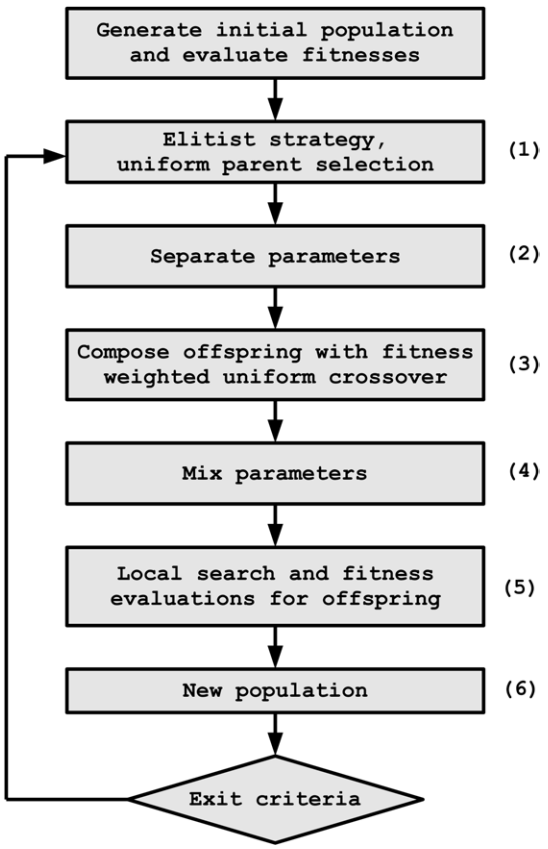

Figure 2. Flowchart of CGA, PCAGA and ICAGA. CGA uses identity transformation in steps (II) and (IV).

(iv) The back transformation of the offspring is obtained by

$$\boldsymbol{X}_{\text{offspring}} = (\boldsymbol{M}\boldsymbol{Y}^{\mathbf{T}}_{\text{offspring}})^{\mathbf{T}}. \tag{9}$$

(v) The fast local random search, which is a simplified simulated annealing algorithm, is used instead of mutation for the children.

(vi) The new trial population is composed from the best individuals of the current population and the offspring.

5. Comparison of genetic algorithms

The scalability of the fitting algorithms was tested with the previously presented nanolaminate model. Each fit was done using a population size of 30 and 25 cycles per fit. The problem was scaled by increasing the number of periods in the layer structure. The results presented in this paper were computed by UltraSPARC IV processors on a Sun Fire 25K server system using Matlab software. The time used for the simulations was 116 h.

Figure 3 shows that CGA has a slightly better performance than PCAGA and ICAGA when the period number is zero. This suggests that linear dependences are not existing between the parameters representing the properties of the same layer but rather between the properties of similar layers. PCAGA and ICAGA have nearly equal performance with the number of periods from 1 to 3. With a higher number of periods, ICAGA

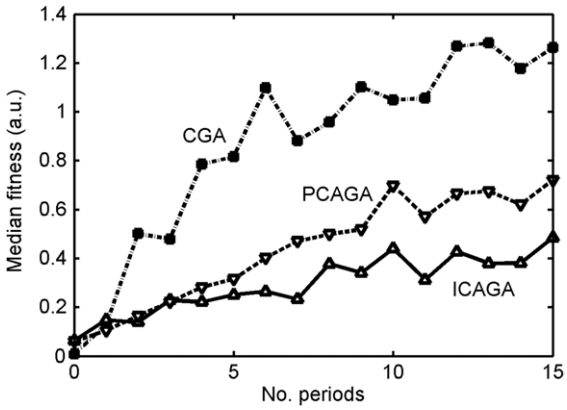

Figure 3. Median fitnesses of CGA, PCAGA and ICAGA when a different number of periods is applied. The median fitnesses are based on 25 fits done for each datapoint.

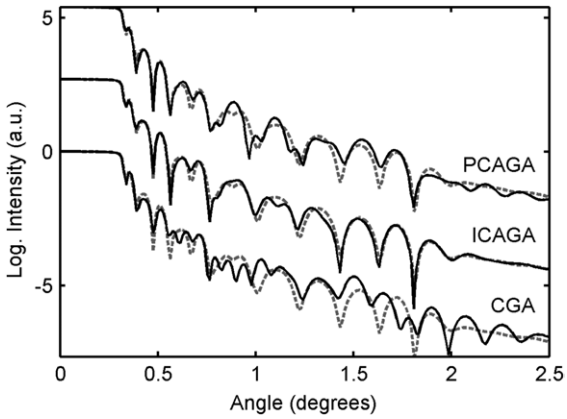

Figure 4. Trial curves of the median of 25 fits in 15 periodic case (black solid line). The target curves are drawn with grey dotted lines. The curves have a vertical offset for clarity.

is outperforming the others. Figure 4 shows the trial curves of the median of 25 fits in 15 periodic case for CGA, PCAGA and ICAGA. The curve fitted by ICAGA is following the target curve clearly better than the curves fitted by other methods.

The median time consumption of one fit used by CGA, PCAGA and ICAGA is shown in figure 5. CGA and PCAGA are linearly scaling but the time consumption of ICAGA increases quadratically. This difference between the novel method and others is a consequence of quadratically growing size of the separation matrix which is iteratively solved by a fixed-point algorithm in EFICA. This sets a practical limit for the problem dimensionality due to the increasing computational requirement but on the other hand, the increasing time consumption is independent of XRR computations. Thus the relative difference in time consumption between the new and other algorithms decreases when the population size or the number of datapoints used in one XRR curve calculation is increased. The additional computational load of EFICA thus decreases relatively when the number of points in XRR curve calculations are multiplied by a factor between three and six which is closer to the number of datapoints in typical real-world XRR measurements.

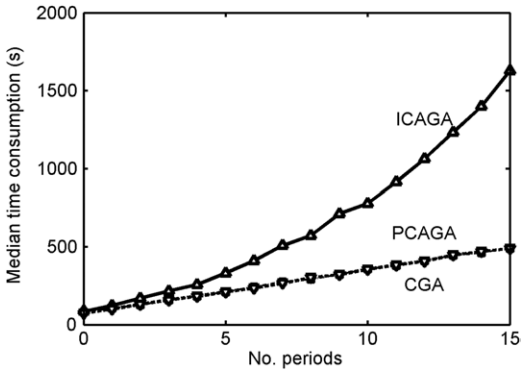

Figure 5. Median time consumption of CGA, PCAGA and ICAGA when a different numbers of periods is applied. The median times are based on 25 fits done for each datapoint.

6. Conclusions

The novel genetic algorithm utilizing ICA, ICAGA, for XRR curve fitting was developed. Genetic linkage between parameters was reduced by the EFICA method during the combination step, which reduces mutual information between parameter distributions by linear transformation. The performance of the new algorithm was studied by fitting trial XRR curves to target curves, which were based on a realistic multilayer model. The capability of GAs to solve XRR fitting problems with an increasing number of parameters was studied with a problem where the number of parameters to be fitted was running from 5 to 140 with a step of 9. CGA performed best when the number of parameters was 5 and with moderately sized problems PCAGA and ICAGA performed equally. With very large problems ICAGA had the best median convergence.

The disadvantage of ICAGA was remarkably greater time consumption in simulations compared with other methods which was due to the need of separation matrix calculation in EFICA. Fortunately this phase is independent of XRR calculations. Note that the time consumption in PCAGA and CGA is mainly determined by the time used in the XRR curve calculations. Time consumption was reduced by a factor between 3 and 6 compared with real-world analysis to decrease several hundred hours of simulation time. Thus the proportion of additional computation time used in ICAGA is decreased significantly in a real-world analysis. Therefore, the major difference between the algorithms was convergence behaviour. The convergence analysis showed that ICAGA has the best capability to extend the analysis based on Parratt's formalism to multiperiodic layer structures.

Acknowledgments

The authors acknowledge the Finnish Agency for Technology and Innovation (TEKES, ALDUS Project) and the Academy of Finland for supporting this work financially. The Finnish IT Center for Science (CSC) is also acknowledged for providing computational resources.

References

[1] Sakurai K and Iida A 1992 *Japan. J. Appl. Phys.* **31** 113–5
[2] Smigiel E and Cornet A 2000 *J. Phys. D: Appl. Phys.* **33** 1757–63

[3] Prudnikov I R, Matyi R J and Deslattes R D 2001 *J. Appl. Phys.* **90** 3338–46
[4] Parratt L G 1954 *Phys. Rev.* **95** 359–80
[5] Nevot L and Croce P 1980 *Rev. Phys. Appl.* **15** 761–80
[6] Wormington M, Panaccione C, Matney K M and Bowen K 1999 *Phil. Trans. R. Soc. Lond.* A **357** 2827–48
[7] Ulyanenkov A, Omote K and Harada J 2000 *Physica* B **283** 237–41
[8] Šimek D, Rafaja D and Kub J 2001 *Mater. Struct.* **8** 16–21
[9] Ulyanenkov A and Sobolewski S 2005 *J. Phys. D: Appl. Phys.* **38** A235–8
[10] Tiilikainen J, Tilli J-M, Bosund V, Mattila M, Hakkarainen T, Airaksinen V-M and Lipsanen H 2007 *J. Phys. D: Appl. Phys.* **40** 215–8
[11] Zhang Q, Allinson N M and Yin H 2000 *2000 IEEE Symp. on Combinations of Evolutionary Computation and Neural Networks* (*San Antonio, TX, USA, 11–13 May 2000*) vol 1 pp 33–6
[12] Takahashi M and Kita H 2001 *Proc. 2001 Congress on Evolutionary Computation* (*Seoul, Korea 27–30 May 2001*) vol 1 pp 643–9
[13] Puurunen R L 2005 *J. Appl. Phys.* **97** 121301
[14] Hyvärinen A, Karhunen J and Oja E 2001 *Independent Component Analysis* (New York: Wiley)
[15] Elam J W and George S M 2003 *Chem. Mater.* **15** 1020–8
[16] Koldovský Z, Tichavský P and Oja E 2006 *IEEE Trans. Neural Netw.* **17** 1265–77