

Publication III

J. Tiilikainen, V. Bosund, M. Mattila, T. Hakkarainen, J. Sormunen and H. Lipsanen, *Fitness function and nonunique solutions in x-ray reflectivity curve fitting: Crosserror between surface roughness and mass density*, Journal of Physics D: Applied Physics **40** (2007) 4259–4263

Reprinted with permission from the publisher

© 2007 IOP Publishing Ltd. (<http://www.iop.org/journals/jphysd>)

Fitness function and nonunique solutions in x-ray reflectivity curve fitting: crosserror between surface roughness and mass density

J Tiilikainen, V Bosund, M Mattila, T Hakkarainen, J Sormunen 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 4 April 2007, in final form 22 May 2007

Published 29 June 2007

Online at stacks.iop.org/JPhysD/40/4259

Abstract

Nonunique solutions of the x-ray reflectivity (XRR) curve fitting problem were studied by modelling layer structures with neural networks and designing a fitness function to handle the nonidealities of measurements. Modelled atomic-layer-deposited aluminium oxide film structures were used in the simulations to calculate XRR curves based on Parratt's formalism. This approach reduced the dimensionality of the parameter space and allowed the use of fitness landscapes in the study of nonunique solutions. Fitness landscapes, where the height in a map represents the fitness value as a function of the process parameters, revealed tracks where the local fitness optima lie. The tracks were projected on the physical parameter space thus allowing the construction of the crosserror equation between weakly determined parameters, i.e. between the mass density and the surface roughness of a layer. The equation gives the minimum error for the other parameters which is a consequence of the nonuniqueness of the solution if noise is present. Furthermore, the existence of a possible unique solution in a certain parameter range was found to be dependent on the layer thickness and the signal-to-noise ratio.

1. Introduction

X-ray reflectivity (XRR) is a fast and inexpensive tool for non-contact thin film metrology. Properties such as film thickness, mass density, and interface/surface roughnesses are obtained by fitting the theoretical curve calculated from Parratt's formalism [1] and by utilizing the Nevot–Croce approximation for roughness [2]. In general, fitting algorithms use a fitness function to obtain the difference between the measured and the calculated curves and the solution is found by minimizing the fitness function [3]. It is known that XRR curve fitting problems have nonunique solutions [4, 5] but the nature and behaviour of nonunique solutions have not been discussed. There is no clear understanding as to how the fitness function should be designed to be resistant against errors in

XRR curves. One way to analyse the selected fitness function on the fitting properties is to use so-called fitness landscapes. The very basic concepts related to fitness landscapes can be found from the written but briefly expressed reference [6]; a fitness landscape is a map where the height represents the fitness as a function of the parameter deviations from the optimal parameters. Therefore, a properly visualized fitness landscape can be used as a direct method to study the robustness of the fitness function and to find parameters which have equal fitness to the known optimal fitness, i.e. nonunique solutions. Unfortunately, this approach has not been actively utilized in XRR curve fitting problems.

In this work a fitness function is designed to be resistant against some typical errors in XRR curves, such as noise, misalignment and improper scaling. This function is used

to study the nonunique solutions with fitness landscapes in a noiseless case. Since the use of fitness landscapes requires the use of known parameters, a realistic atomic-layer-deposited (ALD) aluminium oxide (AlO) layer model is used as a test case. The deviation of parameters from known values is performed in a physical parameter and an ALD process parameter space. The physical parameter space is the input space with dimensions of thickness, mass density and roughness for each layer. The process parameter space is the input space where the most important growth parameters in ALD, the number of growth cycles and the deposition temperature, are used as input values. The deviated process parameters are mapped by neural networks to the physical parameters, wherefrom an XRR curve can be directly computed. Using this approach, we show that the XRR curve fitting problem has tracks in fitness landscapes where the local optima are situated. The tracks are utilized in the construction of the crosserror equation between the mass density and the surface roughness of the realistic layer structure. This equation gives a lower bound for the error in these weakly determined parameters. In section 2, general nonidealities existing in typical XRR data are discussed and a robust fitness function is designed to minimize the effect of nonidealities. In section 3, fitness landscapes as a function of the process parameters are visualized and tracks in these landscapes containing local optima are mapped to the physical parameter space by neural networks.

2. Fitness function

2.1. Preprocessing of data

XRR measurements have several sources of systematic errors, e.g. a sample holder is misaligned, a detector is saturated or the attenuating factor of an applied metal attenuator is improperly determined. In the case of the invalid attenuating factor, the measured curve is vertically shifted. Horizontal and vertical shifts are caused by misalignment which prevents the proper determination of mass density and roughness. In studies of a great number of samples, misalignment due to human error is difficult to avoid and the misalignment can also be a consequence of the limited angular resolution used in a measurement. Figure 1 shows a real fitting case without the preprocessing of the data. The measured curve is not properly normalized and noise is visible in the high angle region. In order to decrease the effect of nonidealities, the following data preprocessing procedure was applied in this work for the target and the trial curve.

- (i) Interpolated data calculated by cubic splines are included in the XRR curves. It was found that the interpolation smooths a fitness landscape which is desirable from the gradient optimization point of view.
- (ii) The intensities are averaged uniformly over five adjacent data points to reduce the effect of noise.
- (iii) Intensity values of more than 50% of the critical point intensity are cut off to reduce the effect of the high intensity region. The curves are shifted horizontally so that they coincide at the cut point. Note that this shift can be done only for the curves having only a very slight horizontal offset. Another way to select the cut point is

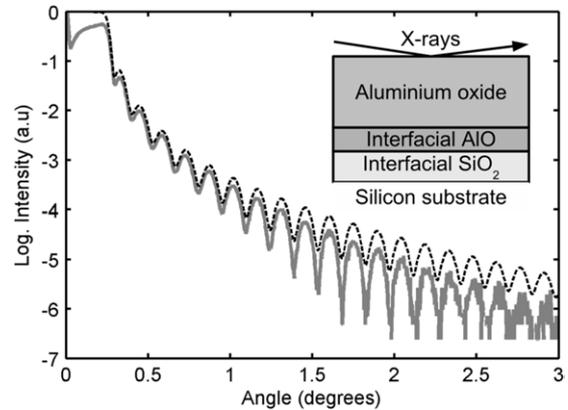


Figure 1. Example of unfinished fitting of the trial curve (black dashed line) to a grey target curve. The trial is a theoretical curve calculated using the layer model shown in the inset. The parameters of the layers are optimized so that the trial and target curves are coincident. The shown target curve is measured from a real world AlO sample. The layer thicknesses are not in scale in the inset.

to use the threshold point used for the attenuator. The threshold point can have a too low value, which causes the important curve area after the critical point to be excluded. This area is used for determining the mass density (or electron density) of the material. Here the selection is made as a compromise between the validity of the data and the sensitivity to mass density although the optimal selection is not known and is therefore somewhat problematic.

- (iv) Both XRR curves are transformed to the logarithmic scale and subsequently scaled so that the integrated intensity is one. This scaling shifts the curves vertically.

2.2. The selection of statistical measure

The selected statistical measure should fulfil the following criteria: (i) it should be fast to compute, (ii) it should attract the trial towards a global optimum as robustly as possible and (iii) it should be as robust as possible against noise. To study the second and the third requirements, fitness landscapes were used as a tool and the following structure for XRR curve calculations was constructed: an aluminium oxide (Al:O, 2:3) layer having a thickness of 40 and 0.55 nm RMS surface roughness with mass density of 3.0 g cm^{-3} was assumed to be on the silicon surface with an interfacial roughness of 0.3 nm. A theoretical XRR target curve based on the model was computed and trial curves were calculated around the given values. The deviating parameters were selected to be the mass density and the surface roughness of the AlO layer.

The properties of several measures such as measures based on entropy, the symmetrized Kullback–Leibler divergence [7] and the Jensen–Rényi divergence [8], were investigated with the model. These measures were tested in their original form and with some modifications with no success. These divergence measures had poor convergence properties in the vicinity of optimal solutions indicating a contradiction to requirement (ii). Several norms having different orders were tested as a fitness measure as well. It was noted that higher order measures are increasingly sensitive to noise

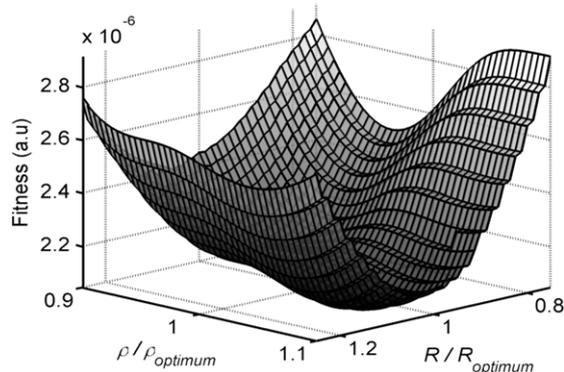


Figure 2. The fitness landscape of a noisy case computed using the measure from [9]. 600 datapoints with four times interpolated data were used in the simulation. R means surface roughness and ρ the mass density. The thickness of the layer and other parameters were fixed in the calculations.

with increasing order but conventional second order moment-based measures satisfied all requirements (i)–(iii). Several modifications of the second order moment based measures were tested and the best two are presented here.

The second best measure was taken from [9]:

$$\text{Fitness} = 1 - [1 + e]^{-1}, \quad (1)$$

where

$$e = \text{RMSE}(x_c, x_m)[1 + r(x_c, x_m)]^{-\alpha}. \quad (2)$$

The constant $\alpha = 3$, x_c is the calculated curve, x_m is the target curve, RMSE means the root mean square error and r is the correlation coefficient of the two arguments. Figure 2 shows the fitness landscape computed using this measure when a uniformly distributed artificial noise with an amplitude of 5×10^{-7} of the maximum intensity was included in the target curve. The landscape exhibited a clear shift of the global minimum when the noise was applied and the case illustrates how clear the shift was even when the second best measure was applied.

A similar test was performed for the $(\chi^2)^{1/2}$ measure. The measure is defined here by

$$(\chi^2)^{1/2} = \left(\sum_{i=1}^n e_i^2 \right)^{1/2} \quad (3)$$

and the error component

$$e_i = |x_{i,m} - x_{i,c}| \cdot x_{i,c}^{-1}, \quad (4)$$

where $x_{i,m}$ is the component of the target curve and $x_{i,c}$ is the component of the trial curve. Note that the relative error in equation (4) is defined with respect to calculated data which is done to prevent the noise in the measurement to distort the shape of the fitness landscape. Also note the use of square root which was selected to slightly linearize a paraboloidal shaped landscape. A paraboloidal landscape has an almost flat fitness region near an optimum, which is not desirable when good selectivity between a good and a better solution is needed. Figure 3 shows the fitness landscape calculated using the $(\chi^2)^{1/2}$ measure with noise. Simulations showed that this measure preserves the minimum fitness valley near the

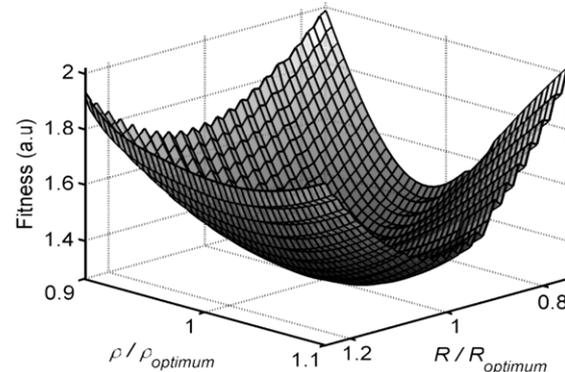


Figure 3. The fitness landscape of a noisy case computed using the $(\chi^2)^{1/2}$ measure. 600 datapoints with four times interpolated data used in the simulation. R means surface roughness and ρ the mass density. The thickness of the layer and other parameters were fixed in the calculations.

optimal value $(\rho_{\text{optimum}}, R_{\text{optimum}})$ although a slight shift of the minimum point can be observed. The $(\chi^2)^{1/2}$ measure was found to be the most robust against errors. All other measures based on the second order moment had no clear minimum in the given range. Similarly, algebraic modifications to obtain a binding limit for the value range of fitness functions failed; robustness was lost since no valley with minimum fitness existed for these modifications.

3. Nonunique solutions

The robustness of a fitness function is important when studying nonunique solutions since it is difficult to estimate the consequences of noise due to several factors affecting it. Here, the $(\chi^2)^{1/2}$ -based fitness function was used to study the behaviour of nonunique solutions with a case study of a realistic model of ALD-grown AIO. Note that the previous model used to study fitness function properties was oversimplified and the thickness in the fitness landscape calculations was fixed. In real world structures simultaneous variations in three physical parameters for each material must be considered. In such a case, the computational time needed to explore the physical parameter space grows exponentially with increasing dimensionality and therefore dimensionality reduction is mandatory. This can be done with feedforward neural networks (NN) as illustrated in figure 4. NNs allowed the simultaneous change in several variables without fixed parameters since the properties of ALD thin films are mainly controlled by two process parameters, by the deposition temperature and the number of cycles [10]. Table 1 summarizes the physical parameters and the neural networks used in the model based on the empirical XRR properties collected from 32 ALD grown AIO samples where these 32 datasets were used to construct the dependence between the process parameters and the physical parameters. The majority of the roughness parameters were set to be constant with realistic values since empirical data provided no trends with these parameters. The details of the construction and the training of the NNs will be reported elsewhere.

The layer model was used for XRR curve calculations in the study of nonunique solutions, and figure 5 shows the

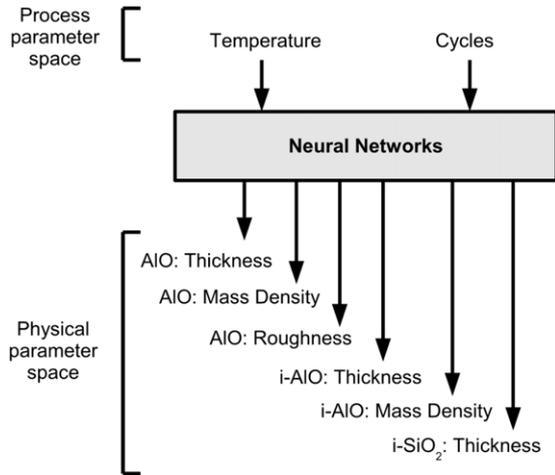


Figure 4. Dimensionality reduction of the problem by neural networks. The two ALD process parameters work as input parameters, which are used to compute six output variables to the physical parameter space. Abbreviation i-AIO and i-SiO₂ means interfacial AIO and interfacial SiO₂, respectively.

Table 1. Parameters used to construct a realistic AIO layer model for XRR curve calculations. NN means a neural network which is used to calculate a certain physical parameter from ALD process parameters, i.e. the number of cycles and the deposition temperature. Abbreviation i-Al₂O₃ and i-SiO₂ means interfacial AIO and interfacial SiO₂, respectively.

Material	Thickness (nm)	Mass density (g cm ⁻³)	Roughness (nm)
Al ₂ O ₃	NN	NN	NN
i-Al ₂ O ₃	NN	NN	0.4
i-SiO ₂	NN	2.6	0.3
Si	∞	2.33	0

flowchart of the applied algorithm used in the study. The fitness landscapes were computed around a given optimal solution as a function of cycles and temperature. Limits used in the calculations were ± 25 cycles and $\pm 10^\circ\text{C}$ which were considered to be realistic upper limits to simulate the nonuniformity of thickness within the substrate area and the inaccurate control of the deposition temperature. Note that film uniformity is slightly ALD machinery design-dependent which causes XRR thin film properties to be also dependent on the location of a wafer inside a reactor and on the nonuniform heat profile. Figure 6 shows a fitness landscape as a function of the process parameters. One can note that the landscape has a track where local optima are located. When noise was included in a target curve, a sharp minimum vanished in earlier fitness landscapes which indicates that noise generally smooths minima. Smoothing of the global optimum region in NN modelling based fitness landscapes was also observed in most cases when the layer thickness was reduced. Thus, it can be concluded that the XRR fitting problem can have a unique solution within some range but in a real world case the existence of a global optimum also depends on the layer thickness and the signal-to-noise ratio. It is worth noting that parameter ranges need to be set carefully to limit the search space to decrease the possible existence of local optima having equal fitness with the global optimum.

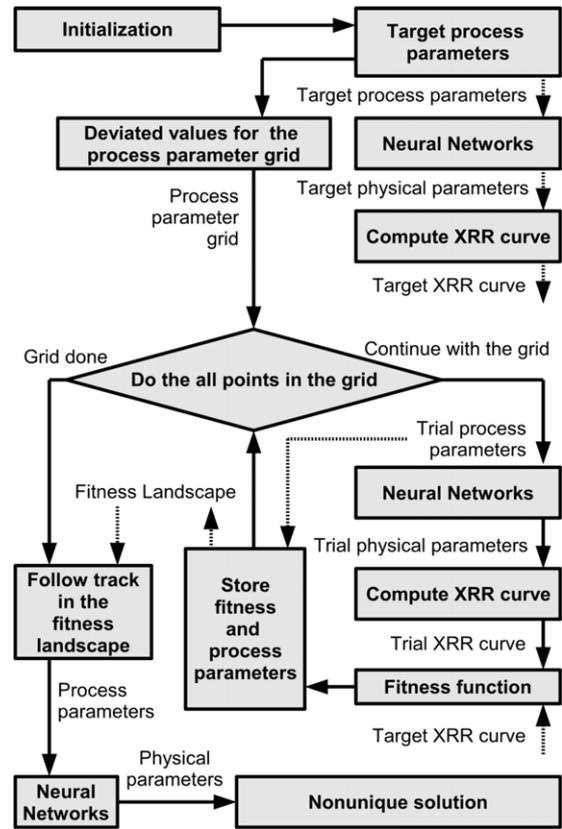


Figure 5. Flowchart of the algorithm used in the study of nonunique solution for the each target process parameter pair. Black solid lines represents the successive steps including dataflow done in the algorithm. Dotted line means dataflow. The physical parameters were calculated using neural networks and missing parameters were set to be constant.

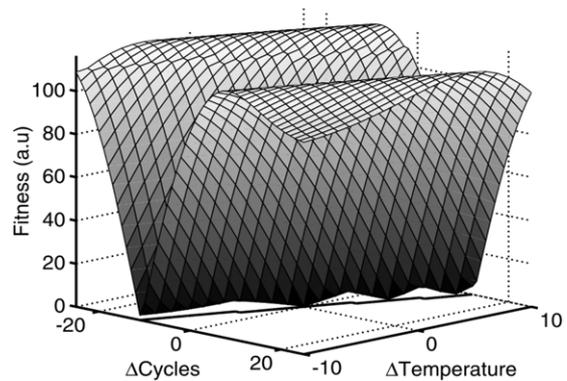


Figure 6. The fitness landscape as a function of the ALD process parameters in the vicinity of the origin. The origin is at temperature 300 °C and 500 cycles in the process parameter space. The black line in the (ΔCycles , $\Delta\text{Temperature}$)-plane is a projection of a track where the local optima lie.

Tracks following local optima were collected from the matrix of fitness landscapes, where global optima were ranging from 100 to 300 °C with steps of 25 °C and from 50 to 500 cycles with steps of 50 cycles. The existence of these tracks in the process parameter space is due to the simultaneous nonlinear change in the physical properties. Figure 7 shows the equivalent physical parameters of the AIO layer which

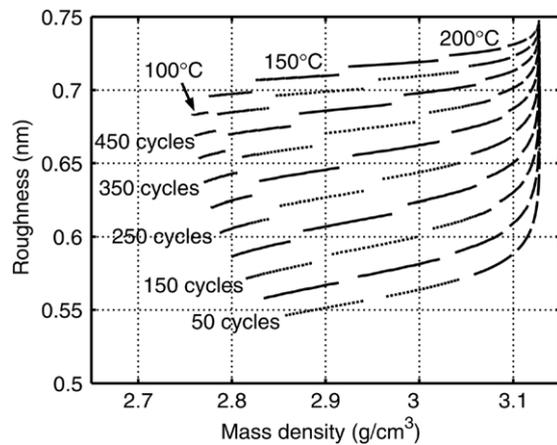


Figure 7. Tracks where local optima lie projected to the (mass density, roughness)-plane. The number of cycles and the deposition temperatures show the optimal solutions in the process parameter space where fitness landscapes were computed.

were calculated from the tracks using previously mentioned neural networks. Between 2.85 and 3.05 g cm^{-3} projections of the tracks are linearly independent of the ALD process parameters. By taking the line marked with 300 cycles and 150°C and approximating it as a linear line, the crosserror between roughness and mass density can be approximated from the slope of the line by the equation

$$\Delta \text{Roughness} \approx 0.11 \times \Delta \text{Mass Density}, \quad (5)$$

where roughness is expressed in nanometres and mass density is in g cm^{-3} . The equation gives a lower limit for the error of surface roughness if the error of mass density is known or estimated. Typically the error of mass density is approximated to be about $\pm 0.1 \text{ kg m}^{-3}$ and thus the minimum error of surface roughness caused by the nonuniqueness of the solution is $\sim 0.011 \text{ nm}$ which translates to $\pm 2.2\%$ error for 0.5 nm roughness. However, a definite error limit for the mass density cannot be set in the light of simulations without *a priori* knowledge. Note that simulation suggests that the mass density can be precisely determined independently of roughness when the layers are simulated with the temperature of 300°C. The result is due to the saturation of mass density above 235°C. Thus, this region is not of interest when considering the crosserror. It is worth mentioning that the shape of lines in figure 7 was preserved even when the artificially generated noise was included in the simulations, thus suggesting the validity of the noiseless simulations with the applied fitness function. The shape of lines was also preserved in all cases when the critical angle region was included to the data and no horizontal shift was applied. A more detailed investigation showed that the global optimum is more distinguishable in some fitness landscapes, but if noise is applied, the location of global optimum can change in the given track. Furthermore, the length of the lines increased when the variations in process parameters were increased. Thus, the beginning and the end of an ambiguous region in mass density and roughness is not clearly defined.

4. Conclusions

Nonunique solutions of an XRR fitting problem were studied and a crosserror equation between roughness and mass density was constructed based on the fitness landscape simulations. Simulations were performed with a fitness function designed and tested to be resistant to several nonidealities present in XRR measurements thus improving the robustness and validity of the obtained results. Simulations were based on neural networks modelling the empirical XRR properties of ALD aluminium oxide layers. The use of the models allowed the reduction of the dimensionality problem and made it possible to visualize the fitness landscape with two ALD process parameters. The computed fitness landscapes revealed tracks where these are local optima. By projecting these tracks onto the physical parameter space, a crosserror equation was constructed for the weakly determined parameters. This equation gives the crosserror between mass density and roughness which sets the minimum error for the second parameter. It was also found that a unique solution can exist within carefully aligned physical limits but it requires a thick layer and a good signal-to-noise ratio in the measurement. The drawback of the present approach is related to the modelling of the XRR properties of AIO which limits the generalization of the simulations to one case. However, the obtained results work as a first approximation in the determination of crosserrors in practical problems when analysing real world ALD layers by the XRR curve fitting. We recommend that the data preprocessing step is thoroughly studied from the point of view of convergence properties and fitting accuracy for further work. Nonunique solutions with other materials are also encouraged to be studied to generalize the results found in this paper. We expect that this work together with subsequent investigations opens the possibility of defining confidence limits for XRR determined properties in the near future.

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] Parratt L G 1954 *Phys. Rev.* **95** 359–80
- [2] Nevot L and Croce P 1980 *Rev. Phys. Appl.* **15** 761–80
- [3] Tiilikainen J, Tili J-M, Bosund V, Mattila M, Hakkarainen T, Airaksinen V-M and Lipsanen H 2007 *J. Phys. D: Appl. Phys.* **40** 215–8
- [4] Reiss G and Lipperheide R 1996 *Phys. Rev. B* **53** 8157–60
- [5] van der Lee A 2000 *Eur. Phys. J. B* **13** 755–63
- [6] McCarthy I P 2004 *Int. J. Oper. Prod. Manag.* **24** 124–50
- [7] Lin J 1991 *IEEE Trans. Inform. Theory* **37** 145–51
- [8] He Y, Hamza B A and Krim H 2003 *IEEE Trans. Signal Process.* **51** 1211–20
- [9] Dane A D, Veldhuis A, de Boer D K G, Leenaers A J G and Buydens L M C 1998 *Physica B* **253** 254–68
- [10] Puurunen R L 2005 *J. Appl. Phys.* **97** 121301