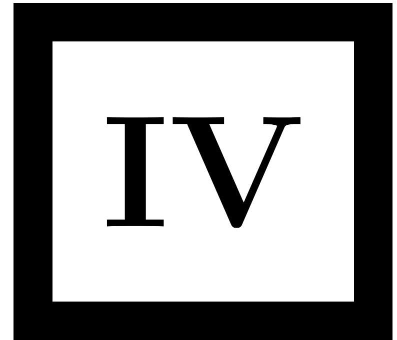

Publication IV

J. Tiilikainen, J.–M. Tilli, V. Bosund, M. Mattila, T. Hakkarainen, J. Sormunen and H. Lipsanen, *Accuracy in x–ray reflectivity analysis*, Journal of Physics D: Applied Physics **40** (2007) 7497–7501

Reprinted with permission from the publisher
© 2007 IOP Publishing Ltd. (http://www.iop.org/journals/jphysd)

J. Phys. D: Appl. Phys. **40** (2007) 7497–7501 doi:10.1088/0022-3727/40/23/037

Accuracy in x-ray reflectivity analysis

J Tiilikainen, J-M Tilli, 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 12 September 2007, in final form 9 October 2007
Published 16 November 2007
Online at stacks.iop.org/JPhysD/40/7497

Abstract
The influence of Poisson noise on the accuracy of x-ray reflectivity analysis is studied with an aluminium oxide (AlO) layer on silicon. A null hypothesis which argues that other than the exact solution gives the best fitness is examined with a statistical p-value test using a significance level of $\alpha = 0.01$. Simulations are performed for a fit instead of a measurement since the exact error caused by noise cannot be determined from the measurement. The p-value is studied by comparing trial curves to 1000 'measurements', each of them including synthetic Poisson noise. Confidence limits for the parameters of Parratt's formalism and the Nevot–Croce approximation are determined in (mass density, surface roughness), (thickness, surface roughness) and (thickness, mass density) planes. The most significant result is that the thickness determination accuracy of AlO is approximately ± 0.09 nm but the accuracy is better for materials having higher mass density. It is also shown that the accuracy of mass density determination can be significantly improved using a suitably designed fitness measure. Although the power of the presented method is demonstrated only in one case, it can be used in any parameter region for a plethora of single layer systems to find the lower limit of the error made in x-ray reflectivity analysis.

1. Introduction

X-ray reflectivity (XRR) is an efficient tool for noncontact thin film metrology. Information such as film thickness, mass density and interfacial roughnesses can be obtained by fitting a theoretical curve based on Parratt's formalism [1] with the Nevot–Croce interface roughness approximation [2] to a measurement. The fitting is carried out by minimizing a fitness function which measures the difference between the target curve and a trial. There are several techniques for doing this procedure efficiently [3–8] but there is no certainty that a found fit is exact if the target curve contains noise. The first attempts to understand the limitations in XRR analysis were studied in [9, 10] but the results were more qualitative than quantitative.

Noise causes an unknown addition to the fitness thus hindering the detection of the exact parameters. In such a case, it is possible that the numerical optimum is not at the same location where the real parameters are. The exact error in analysis due to noise is, however, not possible to determine. Noise causes slight differences between measurements and therefore the effect of noise can be limited with a plethora of measurements but this approach is very impractical. One way to circumvent the problem is to approximate that the fit is the measurement and to use statistical tests to analyse the error of this 'measurement' including synthetic Poisson noise. In this paper, a p-value is used in the validity test of a null hypothesis which says that a solution other than the exact is giving better fitness, i.e. the numerical optimum of the fitness function is elsewhere than at the exactly known location. The confidence limit determination of XRR parameters using the p-value test is realized for a single layer system in two-parameter planes. The background of the work is presented in section 2, the methodology used in confidence limit determination is introduced in section 3 and results in section 4.

2. Background of the work: ϵ-technique

The easiest approach to obtain the error (or confidence) limits for a given solution is to examine the fitness value as a function of parameters. Here it is called an ϵ-technique. The formalized

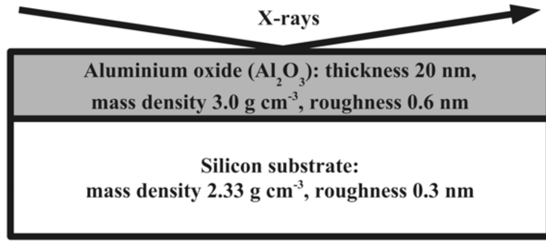

Figure 1. Schematic layer structure used in the modelling of XRR curves. The roughness refer to the surfaces of the layers, i.e. the roughness of the aluminium oxide layer is the surface roughness and the roughness of the silicon substrate is the interfacial roughness.

form of the technique is

$$|F(\boldsymbol{p}_{\text{fit}}) - F(\boldsymbol{p}_{\text{fit}} + \Delta\boldsymbol{p})| < \epsilon, \qquad (1)$$

where $F > 0$ is the function calculating the scalar fitness value between a measurement and a theoretical XRR curve, $\boldsymbol{p}_{\text{fit}}$ is the vector containing XRR parameters of a theoretical fit, a scalar $\epsilon > 0$ is an allowed maximum growth in fitness due to an error $\Delta\boldsymbol{p}$ in the XRR parameters. There are a few strategies to select a suitable ϵ: (1) Set $\epsilon = cF(\boldsymbol{p}_{\text{fit}})$, where constant $c > 0$ sets the error to be determined as a function of $F(\boldsymbol{p}_{\text{fit}})$. (2) Set $\epsilon = c > 0$, i.e. constant. There are several serious problems with this approach: firstly, it is not known which ϵ selection strategy is better. Secondly, the selection of a suitable c depends on factors such as the noise level, the applied fitness measure and on the problem itself. The determination of a suitable c may require using another algorithm. Thirdly, if the variation is done for a single parameter at a time but the error region is nonparaxial, the error is determined inadequately. In fact, this is the case with the problem studied in this work. All in all, the ϵ-technique has serious limitations and therefore a better approach is needed to circumvent the aforementioned problems. In the following section, the methodology for a new approach based on statistical hypothesis testing is introduced, and the approach is applied to a single layer system.

3. Methodology

3.1. Layer structure used in the modelling of XRR curves

X-ray reflectivity is a powerful tool in the characterization of atomic-layer deposited (ALD) materials. ALD is a growth method for thin films where the source gases are introduced separately [11]. One of the most ideal ALD materials is amorphous aluminium oxide (AlO) grown in the trimethylaluminium/water process and this material is easy to characterize with XRR. Therefore XRR properties of ALD AlO were utilized in the simulations of the case study and the simulations were performed using the layer structure shown in figure 1.

3.2. Poisson noise

Synthetic Poisson noise was used in this study to mimic the distribution of discrete photons calculated in continuous time. The number of detected photons obeys the Poisson distribution:

$$P(C) = \frac{\overline{C}^{C}}{C!}\exp(-\overline{C}), \qquad (2)$$

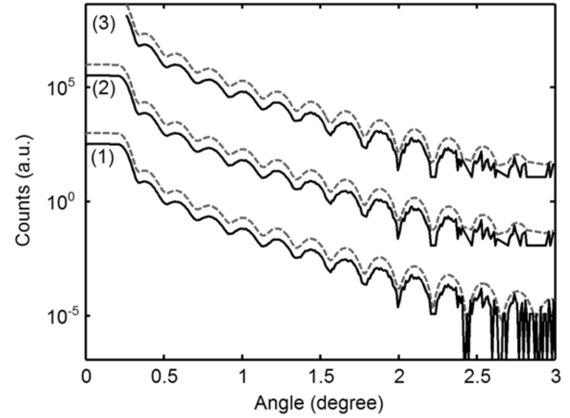

Figure 2. Preprocessing steps in fitness function. The black solid line represents the noisy measurement and the grey dashed line a theoretical curve. (1) The curves after normalization, (2) the curves when the sampling points below one photon count level are discarded and (3) the curves when the angle regions below the critical angle are cut off. The curves have vertical offset for clarity.

where $P(C)$ is the probability distribution of the detected counts C and $\overline{C}$ is the mean number of photon counts. $\overline{C}$ is obtained from the theoretical counts while the number of detected photons C is generated by an algorithm generating random numbers following the Poisson distribution. The number of photon counts $\overline{C} = \overline{I}T$, where $\overline{I}$ is the intensity (photon counts per second) and T is the averaging time. Intensity $\overline{I}$ can be calculated using Parratt's formalism, where $\overline{I} = R\overline{I}_{\text{max}}$, $R = |r|^2$ is the squared modulus of the reflectivity coefficient based on Parratt's equations for the electric field and $\overline{I}_{\text{max}}$ is the measured mean of the intensity with total reflection. In our Philips X'Pert Pro XRR setup, $\overline{I}_{\text{max}}$ was measured to be 2.33×10^6 counts per second and this number was used in the simulations.

3.3. Fitness functions

The purpose of the fitness function is to handle nonidealities of the data and to measure the fitness (goodness-of-fit) between two XRR curves. Recall that the fitness is traditionally calculated between two curves in logarithmic scale but Parratt's formalism and XRR measurement give curves in a linear scale. Therefore, proper design of the fitness function is mandatory to decrease the effect of noise and to select a reasonable balance between the improved sensitivity properties of the measure and decreasing convergence efficiency in fitting algorithms. Here the preprocessing in the fitness functions is done as follows:

(i) The curves are normalized according to the maximum intensity so that $C_{\text{max}} = 1$. Note that the theoretical curve is slightly smoothed with a Gaussian filter to mimic instrumental convolution.
(ii) Sampling points below one photon count level in the original measurement are discarded from both curves.
(iii) The angle region below the critical angle, i.e. sampling points with $C > 0.5C_{\text{max}}$ are cut off.

Figure 2 illustrates how the aforementioned preprocessing steps affect the XRR curve. After the preprocessing steps, the common logarithm of XRR curves is taken. Note that the intensity at the critical angle value is now $\log_{10}(0.5C_{\text{max}})$.

Subsequently, the fitness measure is used to calculate the difference between these curves. The fitness measure, root mean squared relative error (RMSRE), is defined here as

$$\mathrm{RMSRE} = \left(N^{-1} \sum_{i=1}^{N} e_i^2 \right)^{1/2}, \tag{3}$$

where its error component

$$e_i = \left| (x_{i,m} - x_{i,c})(x_{i,c} + x_b)^{-1} \right|. \tag{4}$$

Here $x_{i,m}$ is a data point of the target curve, $x_{i,c}$ is a data point of the trial curve. Bias $x_b > 0$ is a scalar value which tunes the sensitivity of RMSRE to the critical angle region. For root mean squared error (RMSE) the error component is defined as

$$e_i = \left| (x_{i,m} - x_{i,c}) \right|. \tag{5}$$

In our recently published paper [10], a very similar $(\chi^2)^{1/2}$ measure with RMSRE was found to be most robust against noise. The convergence properties of fitting algorithms with this measure were not studied at that time but it came out later in our experiments that the improved accuracy in mass density determination is obtained at the expense of the robustness in convergence properties. In other words, if the mass density of an initial trial differs too much from the optimum, the fitting problem corresponds to a 'needle in a haystack' problem since the invalid value at the critical angle dominates via the denominator in the RMSRE measure. Thus, a poor initial guess in the mass density torpedoes the convergence of solutions having otherwise well-defined parameters. To circumvent this problem, the bias term x_b was included in the RMSRE measure. In the following section, it is shown how the tuning of the bias term affects the confidence limits in (mass density, roughness)-plane.

3.4. Determination of confidence limits

There are four parameters to be optimized in a single layer case: layer thickness, mass density of the layer, surface roughness and interfacial roughness between the layer and the substrate. Typically the mass density of the substrate is known and therefore it was set to be constant.

A preliminary test was conducted to determine which parameters have a major contribution to the fitness of a solution. Fitness values were computed between noiseless trials and the noiseless target curve. The target curve was based on the previously introduced model presented in figure 1. Trials were generated at every point of a four dimensional grid around the target curve with the grid dimensions of layer thickness, layer mass density, surface roughness and interfacial roughness.

It was found that the interfacial roughness between the AlO layer and the substrate affects very weakly the fitness compared with the other parameters. Therefore, the most interesting region depended on the three remaining parameters and formed a three-dimensional object. The determination of its surface is a computationally intensive operation but if the edge of the object is originally computed as a projection onto a two-dimensional plane, the computational load reduces remarkably.

After the preliminary study, a statistical approach was used to determine confidence limits for the case study. Since it is impossible to find out the exact properties of the layer structure from the measurement, a fit was used to approximate the measurement. The exact parameter set taken from the fit was denoted as $\boldsymbol{p} = [t, \rho, R]$, where t is the layer thickness, ρ is the mass density and R is the surface roughness of the layer. A noiseless XRR curve was generated and fitness was calculated between the curve without and with synthetic Poisson noise. This step was repeated $N = 1000$ times, the noisy curves were saved and fitness values were saved in a vector component F_i^{exact}, where $i = 1, 2, 3 \ldots N$. These values were used to determine whether the null hypothesis

$\boldsymbol{H_0}$: *Other than the exact solution gives the best fitness*

is true or not. Here the fitness is calculated between a trial curve and some noisy XRR curve calculated from the exact parameters. Note here that the hypothesis argues that an arbitrarily selected XRR parameter set has better fitness than the exact solution. If this hypothesis is rejected, the exact solution gives better fitness than a trial, otherwise there is not enough evidence against the hypothesis. Strictly speaking, accepting the hypothesis does not mean that the hypothesis is true. The validity determination of the hypothesis is based on a significance level α; if the p-value giving the probability of the null hypothesis with some XRR parameters is less than the given α, the hypothesis is invalid at that location. In that sense, this approach requires testing the validity of the hypothesis at every point of a two–dimensional search space but is impractical due to the huge number of computations required. To reduce computation time, the monotonic behaviour of p-value as a function of distance between trial parameters and the exact solution was assumed. Thus one can use a search algorithm to find such a trial solution in a direction where the condition $p = \alpha$ is fulfilled. The determination of this confidence limit for XRR parameters was done here as follows:

(i) Select the plane, for instance (mass density, surface roughness) plane, where to determine confidence limits.

(ii) Define a search direction $\boldsymbol{d}(\theta)$, where $\theta \in [0^\circ, 360^\circ]$, for instance:

$$\boldsymbol{d}(\theta) = \boldsymbol{p} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \cos(\theta) & 0 \\ 0 & 0 & \sin(\theta) \end{bmatrix} \tag{6}$$

(a) For each θ, search $p(r) \approx \alpha$:

1. Calculate an XRR curve with the parameters $\boldsymbol{p} + r\boldsymbol{d}(\theta)$.
2. Calculate its fitness with aforementioned N exact noisy curves. Save the fitness values to a vector component F_i, where $i = 1, 2, 3 \ldots N$.
3. Calculate

$$p = \frac{1}{N} \sum_{i=1}^{N} (F_i < F_i^{\mathrm{exact}}). \tag{7}$$

The condition $(F_i < F_i^{\mathrm{exact}})$ gives one if satisfied, otherwise zero.

4. If $p \approx \alpha$, save $\boldsymbol{p}_{\mathrm{contour}}(\theta) = \boldsymbol{p} + r\boldsymbol{d}(\theta)$.

(iii) Draw closed contour $\boldsymbol{p}_{\mathrm{contour}}$ in the selected plane.

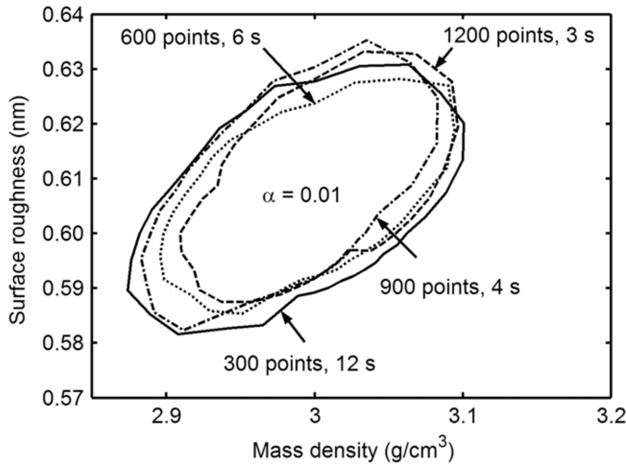

Figure 3. Confidence limits of the RMSE measure with $\alpha = 0.01$ when the measurement time is constant but the averaging time is varied.

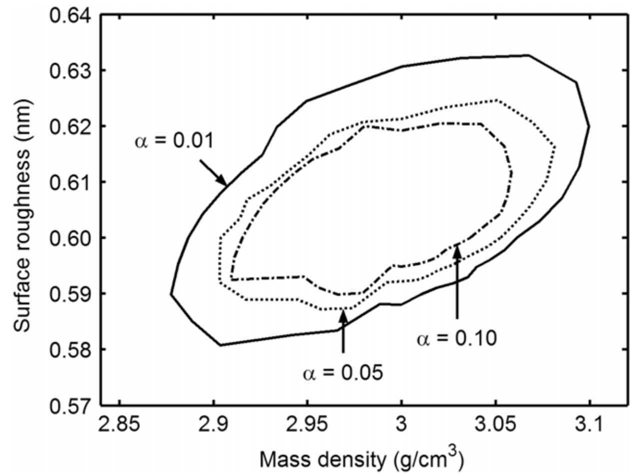

Figure 4. Confidence limits of the RMSE measure with $\alpha = 0.01$, 0.05 and 0.1.

Using the assumption of the monotonous p-value, one can separate rejection (outside) and acceptance (inside) regions, and the sizes and dimensions of these regions indicate rapidly the lower limit for the error of the exact solution. Note that this lower limit is the fundamental limitation of XRR analysis due to Poisson noise and cannot be circumvented. Naturally the real error, which can be a consequence of several factors including nonoptimal fit, can be greater.

4. Results and discussion

The analysis of confidence limits with the case study may depend on the averaging time used in the measurements, i.e. averaging time increases photon count level thus affecting Poisson noise distribution. The trade-off between the number of sampling points versus the averaging time was studied by keeping the total 'measurement' time constant. Simulation mimicking a measurement was performed using 300, 600, 900 and 1200 sampling points with 12 s, 6 s, 4 s and 3 s averaging times, respectively. Figure 3 shows confidence limits as a function of the number of sampling points in XRR curves. One can note that the region with 300 points seems to be the largest while the others are slightly smaller. The regions with 900 and 1200 points seem to be the smallest thus suggesting that it is beneficial to increase the number of points in a measurement to a certain extent rather than to increase the averaging time. However, since the differences in the sizes of the regions were small, 300 points and 12 s averaging time were used in the subsequent simulations due to shorter computation time.

Figure 4 shows the confidence limits as a function of significance level α. One can note that the region inside the limits is not significantly reduced when α is increased from 0.05 to 0.1. Since the confidence limits vary slightly between simulations, these simulations suggest that the position of limits are not changing sufficiently at high significance levels and thus the use of $\alpha = 0.01$ is preferred.

The accuracy of analysis is fitness measure dependent as mentioned earlier. Figure 5 shows that the confidence region in the mass density dimension decreases with decreasing x_b in RMSRE. In other words, the smaller the region size, the better the accuracy that can be obtained from the analysis

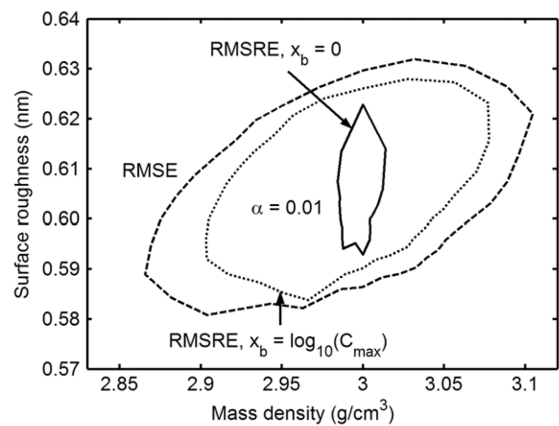

Figure 5. Confidence limits of the RMSE and RMSRE measures with $\alpha = 0.01$.

which proves the earlier claim describing the qualitative difference between RMSRE and RMSE measures. Note that the principal axes of the confidence regions are not paraxial and the directions of the axes are dependent on x_b. This result suggests that the error limit determination based on the one-parameter ϵ-technique does not provide correct results. Despite the significantly improved sensitivity of RMSRE to mass density, the bias term x_b has unknown consequences on the convergence properties and therefore RMSE was selected for the major part of simulations due to its simplicity.

One of the most interesting questions in XRR analysis is the accuracy of the thickness determination. Figure 6 shows that increasing layer mass density improves the accuracy of layer thickness and surface roughness determination. It can be observed that the thickness determination accuracy of AlO is approximately ± 0.09 nm with the confidence level of $\alpha = 0.01$. Although the mass density of the ALD AlO layer is rather close to 3.00 g cm^{-3}, the results show that the analysis accuracy can be improved if the mass density, i.e. electron contrast between the layer and the substrate, can be tailored.

Whereas mass density clearly affects the thickness accuracy, surface roughness has no clear impact on it as seen from figure 7. Surface roughness, however, affects the mass density accuracy but the relationship between

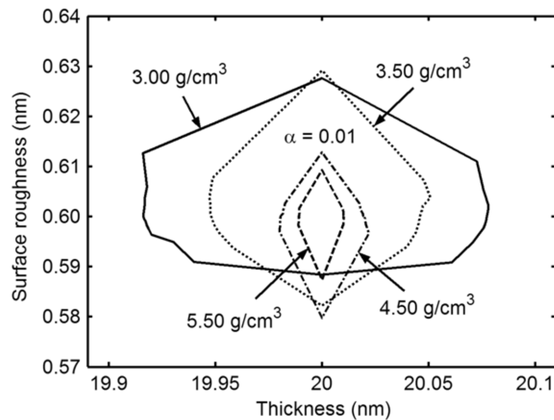

Figure 6. Confidence limits of the RMSE measure with $\alpha = 0.01$ as a function of mass density.

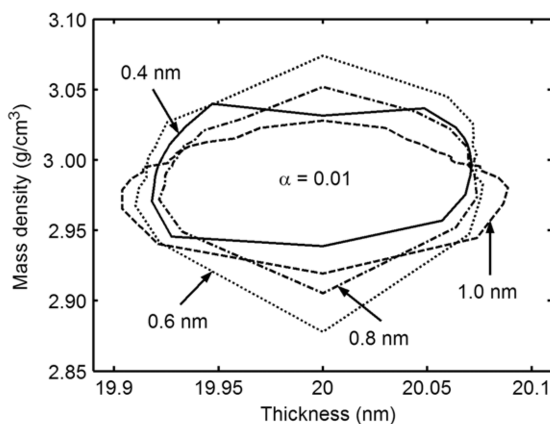

Figure 7. Confidence limits of the RMSE measure with $\alpha = 0.01$ as the function of surface roughness.

the accuracy of mass density and surface roughness seems not to be straightforward.

5. Conclusions

The influence of Poisson noise on the accuracy of XRR analysis was studied with a single layer system case. A hypothesis which argued that other than the exact solution gives the best fitness was examined with the p-value test. Confidence limits for XRR parameters at a certain significance level α were determined in (mass density, surface roughness), (thickness, surface roughness) and (thickness, mass density) planes. The confidence limits were separating rejection (outside) and acceptance (inside) regions of the hypothesis. Note that the applied statistical test says that the hypothesis is incorrect in the rejection region with the given significance level but in the acceptance region there was no enough evidence to reject the hypothesis. In other words, in the rejection region the exact solution has better fitness than the trial solution while in the acceptance region this is not sure. Recall here that the fitness plays an important role in fitting; fitting algorithms try to optimize the fitness and the quality of the solution is determined solely from the fitness value. Therefore an ambiguity in a solution is possible in the acceptance region and this is a consequence of Poisson noise at a certain time which sets the fundamental limit for the accuracy in XRR. Therefore the only way to improve the accuracy in conventional XRR is to increase the number of detected photons by using longer measurement time or higher intensities. As an application of the presented method, the thickness determination accuracy was found to be approximately ± 0.09 nm for a 20 nm thick layer with a significance level of $\alpha = 0.01$ but the accuracy is better with materials having greater mass density. On the other hand, the mass density accuracy was found to be dependent on the applied fitness measure and it was shown that the accuracy can be significantly improved with the suitable selection of fitness measure. Difficulties with the convergence properties when using the sensitive fitness measure, however, limit its applicability. Therefore we recommend that the sensitivity and convergence properties of the fitness measure are thoroughly studied in the future to fully exploit the potential of XRR analysis of thin layers.

Acknowledgments

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

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] Wormington M, Panaccione C, Matney K M and Bowen K 1999 *Phil. Trans. R. Soc. Lond.* A **357** 2827–48
[4] Ulyanenkov A, Omote K and Harada J 2000 *Physica* B **283** 237–41
[5] Šimek D, Rafaja D and Kub J 2001 *Mater. Struct.* **8** 16–21
[6] Ulyanenkov A and Sobolewski S 2005 *J. Phys. D: Appl. Phys.* **38** A235–8
[7] 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
[8] Tiilikainen J, Bosund V, Tilli J-M, Sormunen J. Mattila M, Hakkarainen T and Lipsanen H 2007 *J. Phys. D: Appl. Phys.* **40** 6000–4
[9] Luokkala B B, Garoff S and Suter R M 2000 *Phys. Rev.* E **62** 2405–15
[10] Tiilikainen J, Bosund V, Mattila M, Hakkarainen T, Sormunen J and Lipsanen H 2007 *J. Phys. D: Appl. Phys.* **40** 4259–63
[11] Puurunen R L 2005 *J. Appl. Phys.* **97** 121301