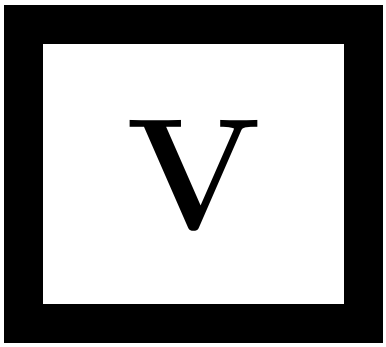

Publication V

J. Tiilikainen, M. Mattila, T. Hakkarainen and H. Lipsanen, *Novel method for error limit determination in x–ray reflectivity analysis*, Journal of Physics D: Applied Physics **41** (2008) 115302

IOP PUBLISHING JOURNAL OF PHYSICS D: APPLIED PHYSICS

J. Phys. D: Appl. Phys. **41** (2008) 115302 (8pp) doi:10.1088/0022-3727/41/11/115302

# Novel method for error limit determination in x-ray reflectivity analysis

**J Tiilikainen, M Mattila, T Hakkarainen and H Lipsanen**

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

E-mail: jouni.tiilikainen@tkk.fi

Received 27 February 2008, in final form 7 April 2008
Published 8 May 2008
Online at stacks.iop.org/JPhysD/41/115302

**Abstract**

A novel error limit determination method for x-ray reflectivity (XRR) analysis is developed and applied to data measured from atomic-layer-deposited aluminium oxide on silicon. The analysis here is based on Parratt's formalism and on a fitness defined as a mean-squared error between a measurement and a fit in logarithmic scales. The mathematically derived upper bound for an error uses a trick which divides the fitness into two parts. The divided original fitness equals the fitness between a measurement and a numerically optimal but unknown fit (the first part) plus the fitness between the unknown optimal fit and the known original fit (the second part). In practical error determination, the fitness in the first part is the fitness of noise and it is approximated using a separate simulation and in the second part, the unknown optimal fit is considered as a variable to be optimized. An efficient implementation is presented for the error determination and the determined parameters were $42.4 \pm 0.12$ nm (0.3%), $3.15 \pm 0.11$ g cm$^{-3}$ (3.5%) and $0.80 \pm 0.06$ nm (7.5%) for the thickness, the mass density and the surface roughness, respectively. Although the formalized error may need some fine tuning as future work since it gives an asymptotic estimate, it still gave reasonable results in the case of systematic error caused by nonideal fit.

## 1. Introduction

x-ray reflectivity (XRR) is a powerful noncontact technique for the nanoscale metrology of thin films. XRR is similar to one-wavelength optical reflectometry but the measurements are typically carried out using the wavelength of Cu K$\alpha$ which provides very high sensitivity to the structural properties of a layer structure. These properties can be determined by fitting a theoretical curve based on Parratt's formalism [1] with Nevot–Croce roughness [2] to a measurement. Several studies concentrating on this challenging inversion topic have been published within the last decade [3–8] but a formally correct and objective method providing an error in the determined parameters after a fitting procedure is still missing.

Luokkala *et al* [9] studied this problem with surfactant monolayers and observed that a single fit can yield misleading structural parameters and multiple statistically equivalent fits are required to produce a more clear picture of the parameters. Although the method is simple to implement and therefore interesting, it may be computationally too expensive to find an adequate number of statistically equivalent fits to be used in error determination. Instead of a statistical brute force approach, we have previously studied a deterministic approach to explain the crosserror between the surface roughness and mass density of a single layer system [11]. In that work, the sensitivity of XRR parameters to data were analysed using deposition process modelling as a tool and the applied approach allowed one to construct the first approximation for the error made between the mass density and the surface roughness. However, the approach always requires the modelling of the deposition process. The need for a more generalized, computationally inexpensive, objective and reliable error limit estimation was the driving force for the second paper on this topic [12]. In that work we introduced a method in which the rejection of a null hypothesis arguing that any other than the exact solution gives the best fitness (goodness-of-fit) is studied using a statistical $p$-value test. It was found that the Poisson noise is an important factor which can cause a numerical

 

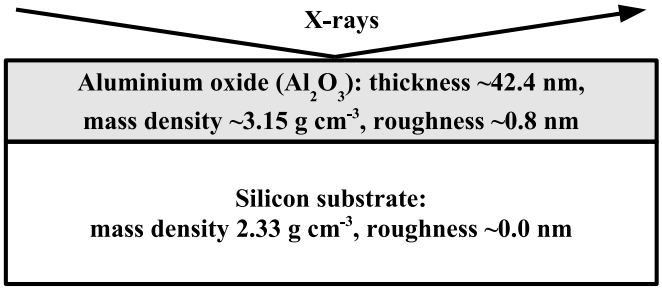


**Figure 1.** Layer model used in the calculation of XRR curves using Parratt's formalism. All the parameters except the mass density of silicon substrate are fitted.

optimum being elsewhere than at the exact solution. Although the presented method gives an objective way to understand and define confidence boundaries for the determined parameters, a major limitation comes from the assumption that a fit with included synthetic Poisson noise is approximately equal to the measurement. Strictly speaking this is not true. Therefore, a more accurate method taking into account the consequences of a nonideal fit and giving a reasonable error in XRR parameters is still required.

In this work, a novel objective method using a formally derived error in XRR analysis is implemented. The derivation of the error uses a simple trick which divides an applied fitness measure into two parts, where the first part measures the error between a measurement and a numerically optimal but unknown fit and the second part measures the error between the unknown optimal fit and a fit given by a fitting algorithm. With this approach, the error determination reduces to a simple optimization problem where the first error term can be estimated from a separate simulation while the second term is optimized so that the artificially separated fitness measure equals the fitness between the measurement and the fit. In section 2, the methodology including the applied layer model, the generation of the synthetic Poisson noise and the used fitness function are presented. In section 3, the old fashioned method with its disadvantages is discussed, the theory behind the novel method is introduced and one of its possible implementations is presented. Finally, the error limits determined for a measurement and the applied layer structure are shown in section 4.

## 2. Methodology

### *2.1. Layer structure used in the modelling of XRR curves*

In atomic layer deposition (ALD) the source gases are introduced sequentially which causes the growth to be chemically self-limited. The self-limiting growth allows the digital thickness control down to a subnanometre range with extreme conformality and therefore ALD grown materials are very ideal to be measured with XRR. One of the most idealistic ALD materials is amorphous aluminium oxide (AlO) grown in the trimethylaluminium/water process [13] which has a reasonable electron density contrast with the silicon substrate and roughness typically well below 1 nm in a large horizontal range. Due to these benefits of ALD AlO, it was used as a case study in this work. The layer structure and material system properties are shown in figure 1. The parameters are determined by fitting a theoretical curve $\overline{x}_{\text{fit}}$ to a measurement $x_{\text{meas}}$. Note that the roughness of the substrate is negligible after a Levenberg–Marquadt based fitting although the typical roughness is a couple of ångströms. This can be a consequence of the nominally thick AlO layer which hinders the determination accuracy of the intentionally small roughness value.

### *2.2. Poisson noise*

In XRR measurements, the number of detected photon counts follows the Poisson distribution:

$$\text{Prob}(C) = \frac{\overline{C}^C}{C!}\exp(-\overline{C}), \tag{1}$$

where $\text{Prob}(C)$ is the probability distribution of the detected counts $C$ and $\overline{C}$ is the expected number of photons. When mimicking synthetic Poisson noise in XRR, the expected number $\overline{C} = C_{\max}\overline{x}_i$, where the $i$th point $\overline{x}_i \in ]0, 1]$ represents the squared modulus of the reflectivity coefficient for the electric field based on Parratt's formalism and $C_{\max}$ is the maximum number of detected photon counts below the critical angle for total reflection. Using these notations, the function generating random numbers from the Poisson distribution is $x = C_{\max}^{-1}\text{Poisson}(C_{\max}\overline{x})$ where $x = (x_1, x_2, \ldots, x_i, \ldots x_N)$.

### *2.3. Fitness function*

Before the calculation of the fitness (goodness-of-fit) between two curves, preprocessing of data must be carried out. Here the preprocessing involves the following:

(i) The curves are normalized according to the maximum intensity so that $\text{Max}(x) = 1$ for both curves.
(ii) Points having values less than $C_{\max}^{-1}$ are discarded from both curves if the logarithmic functions are applied.
(iii) The total external reflection region is cut off from both curves since the region below does not agree with the reflectivity coefficients computed using Parratt's formalism without geometrical modelling. Some researchers cut the curves from the critical angle region, i.e. where $x > 0.5\text{Max}(x)$, but preliminary analysis showed that this approach gives practically equal results with the prementioned approach.

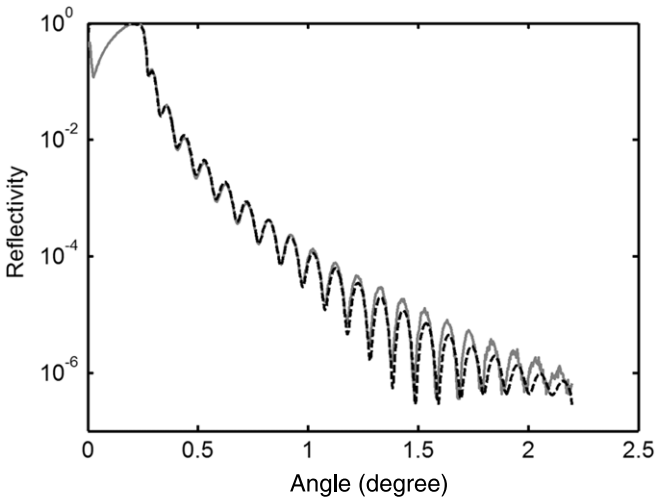


**Figure 2.** Measured (grey solid line) and fitted (black dashed line) XRR curves of an ALD AlO sample. The black curve is fitted using the $\chi^2$ measure.

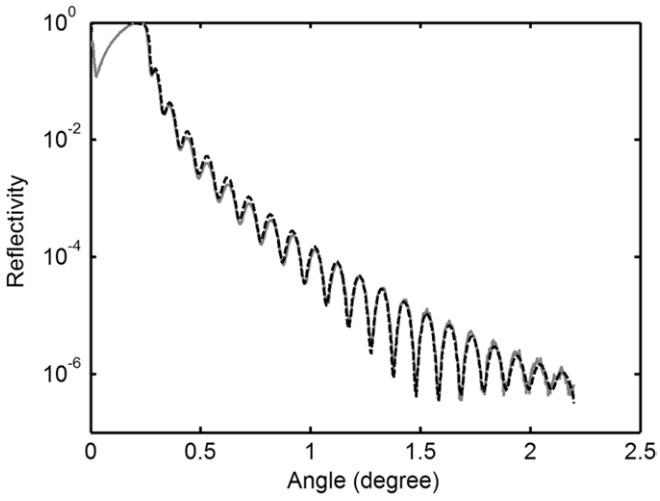


**Figure 3.** Measured (grey solid line) and fitted (black dashed line) XRR curves of an ALD AlO sample. The black curve is fitted using the $F$ measure.

For the preprocessed curves, the selection of fitness measure is important. One can use a measure denoted here as

$$\chi^2(x_{\mathrm{meas}}, \overline{x}_{\mathrm{fit}}) = \sum_{i=1}^{N} \frac{(x_{i,\mathrm{meas}} - \overline{x}_{i,\mathrm{fit}})^2}{\overline{x}_{i,\mathrm{fit}}}, \tag{2}$$

where $x_{i,\mathrm{meas}}$ is the $i$th point of the measured curve $x_{\mathrm{meas}}$ and $\overline{x}_{i,\mathrm{fit}}$ is the $i$th point of a noiseless theoretical XRR curve $\overline{x}_{\mathrm{fit}}$ is calculated using the function $\overline{x} = \mathrm{calcXRRcurve}(p)$ where $p$ is a parameter set used for the XRR curve calculation. For the further analysis, we emphasize here that $\overline{x}$ denotes a theoretical curve without noise and $x$ refers to a curve containing noise. Another fitness measure is here defined as

$$F(x_{\mathrm{meas}}, \overline{x}_{\mathrm{fit}}) = N^{-1} \sum_{i=1}^{N} (\log x_{i,\mathrm{meas}} - \log \overline{x}_{i,\mathrm{fit}})^2 \tag{3}$$

$$= N^{-1} \parallel \log x_{\mathrm{meas}} - \log \overline{x}_{\mathrm{fit}} \parallel_2^2 . \tag{4}$$

Figures 2 and 3 show the difference between XRR curves fitted by different measures. The same initial trial curve was fitted by the Levenberg–Marquadt method to the measurement using $\chi^2$ and $F$ measures. The curve fitted by $\chi^2$ has almost perfect fit in the angle range below 1.25° whereas the curve fitted by $F$ has sufficient fit below 0.9° but almost perfect after. Thus, one can conclude that $\chi^2$ has better sensitivity into the beginning of the curve and therefore the mass density can be determined more accurately with $\chi^2$ than with $F$. However, the curve fitted by $\chi^2$ suffers from clear inaccuracy problem in the tail of the curve. Such a problem is not seen with the curve fitted using the $F$ measure and it has better agreement visually although it has slight but not as serious inaccuracy as $\chi^2$. Recall that the visual agreement is typically used in scientific papers to justify the correctness of results and therefore $F$ has practical significance. Based on the better visual correspondence obtained using $F$, the further study is concentrated on this logarithmic expression. The analysis based on $\chi^2$ is done in comparison purposes although the focus is on the determination of error limits using $F$.

## 3. Methods for error limit determination

### *3.1. Classical approach for the error limit determination*

Commonly utilized technique for the error limit determination, here denoted as $\epsilon$ technique is defined as

$$F[x_{\mathrm{meas}}, \overline{x}(p_{\mathrm{fit}} + \Delta p)] < F_{\mathrm{min}} + \epsilon, \tag{5}$$

where $F > 0$ is the fitness function calculating the fitness value between a measurement $x_{\mathrm{meas}}$ and a theoretical XRR curve $\overline{x}(p)$ as a function of the parameter set $p$, $F_{\mathrm{min}}$ is the minimum fitness and $\epsilon$ is a confidence limit. If $F_{\mathrm{min}}$ and $\epsilon$ are defined, the easiest approach is to variate one parameter from $p$ at a time. However, this so-called monovariate approach estimates incorrectly the error if correlation exists between the parameters [12]. A multivariate technique can circumvent this problem. For instance, one can use the eigenvectors of a Hessian matrix to find such directions which do not have these interparameter correlations in the determination of proper error limits. In practice, some problems still exist with this approach. In this $\epsilon$ technique

- one assumes that [9]

$$F_{\mathrm{min}} = F_{\mathrm{fit}} = F[x_{\mathrm{meas}}, \overline{x}(p_{\mathrm{fit}})], \tag{6}$$

  which is not guaranteed. Recall that the convergence criteria of fitting algorithms are rather heuristical, i.e. the algorithms do not know the exact value for the minimum of the fitness.
- $\epsilon$ can be determined based on degrees of freedom [9] or on the basis of the $p$-test [10]. One should note that these approaches are not taking into account systematic errors caused by a nonideal fit.

Thus, it can be concluded that the $\epsilon$ technique has some significant limitations and new approaches for the error limit determination are thus welcome. A novel approach having none of the above mentioned problems is introduced in the following section.

### 3.2. Novel method

*3.2.1. Theory of the method.* Consider the case when the fit is perfectly optimized, i.e. a theoretical curve is fitted to a measurement in such a way that the fitness value cannot be minimized more. In that case, the minimum possible fitness caused by noise

$$F_{\text{noise}} = F(x_{\text{meas}}, \overline{x}_{\text{opt}}), \\ = N^{-1} \parallel \log x_{\text{meas}} - \log \overline{x}_{\text{opt}} \parallel_2^2, \quad (7)$$

where $\overline{x}_{\text{opt}}$ is a numerically optimal XRR curve to minimize $F$ with $x_{\text{meas}}$. Note that this is an important special case. In the fitting procedure one minimizes the fitness measure as a function of the parameter set $p$. If the target curve is noiseless, one knows the minimum fitness value which is zero. Therefore, one can run the fitting algorithm until zero fitness is obtained or the fitness value is below an initially set constant. This is not possible when the target curve contains noise. The fitness is greater than zero but its exact value is unknown. However, it will be shown later that the distribution of $F_{\text{noise}}$ can be simulated in order to get an estimate for the minimum of $F_{\text{noise}}$. Before that, the error term $\log x_{\text{meas}} - \log \overline{x}_{\text{opt}}$ in equation (7) can be utilized in an analytic fashion: let us denote fitness

$$F(x_{\text{meas}}, \overline{x}_{\text{fit}}) = N^{-1} \parallel \log x_{\text{meas}} - \log \overline{x}_{\text{fit}} \parallel_2^2 \\ = N^{-1} \parallel (\log x_{\text{meas}} - \log \overline{x}_{\text{opt}}) \\ + (\log \overline{x}_{\text{opt}} - \log \overline{x}_{\text{fit}}) \parallel_2^2 . \quad (8)$$

By shortening the error term of $F_{\text{noise}}$ as $e_{\text{min}} = \log x_{\text{meas}} - \log \overline{x}_{\text{opt}}$ and writing the residual term $e_{\text{res}} = \log \overline{x}_{\text{opt}} - \log \overline{x}_{\text{fit}}$ between the noiseless curves, one obtains for sufficiently large $N$ with good signal-to-noise ratio (see the appendix)

$$F(x_{\text{meas}}, \overline{x}_{\text{fit}}) = N^{-1} \parallel e_{\text{min}} + e_{\text{res}} \parallel_2^2 \\ \geqslant N^{-1} \parallel e_{\text{min}} \parallel_2^2 + N^{-1} \parallel e_{\text{res}} \parallel_2^2 . \quad (9)$$

Since $F(x_{\text{meas}}, \overline{x}_{\text{opt}}) = N^{-1} \parallel e_{\text{min}} \parallel_2^2$ and $F(\overline{x}_{\text{opt}}, \overline{x}_{\text{fit}}) = N^{-1} \parallel e_{\text{res}} \parallel_2^2$, the upper bound for the error

$$F(\overline{x}_{\text{opt}}, \overline{x}_{\text{fit}}) \leqslant F(x_{\text{meas}}, \overline{x}_{\text{fit}}) \\ - F(x_{\text{meas}}, \overline{x}_{\text{opt}}), \quad (10)$$

where the values of $\overline{x}_{\text{opt}}$ are to be searched. Note that $F(x_{\text{meas}}, \overline{x}_{\text{fit}})$ is known after the fitting procedure and $F(x_{\text{meas}}, \overline{x}_{\text{opt}}) \approx F(x_{\text{opt}}, \overline{x}_{\text{opt}})$ can be approximated by simulations. Typically after the fitting process $\overline{x}_{\text{fit}} \approx \overline{x}_{\text{opt}}$. Therefore $F(x_{\text{opt}}, \overline{x}_{\text{opt}}) \approx F(x_{\text{fit}}, \overline{x}_{\text{fit}})$ and

$$F(\overline{x}_{\text{opt}}, \overline{x}_{\text{fit}}) \leqslant F(x_{\text{meas}}, \overline{x}_{\text{fit}}) \\ - F(x_{\text{fit}}, \overline{x}_{\text{fit}}). \quad (11)$$

The values of $F(x_{\text{fit}}, \overline{x}_{\text{fit}})$ can be obtained by simulations with the following steps:

(i) Set $i = 1$.
(ii) Set a new seed number for the random number generator.
(iii) Add synthetic Poisson noise to the noiseless XRR curve according to the equation

$$x_{\text{fit}} = C_{\text{max}}^{-1} \text{Poisson}(C_{\text{max}} \overline{x}_{\text{fit}}) \quad (12)$$

with the given seed.
(iv) Compute $g(i) = F(x_{\text{fit}}, \overline{x}_{\text{fit}})$.

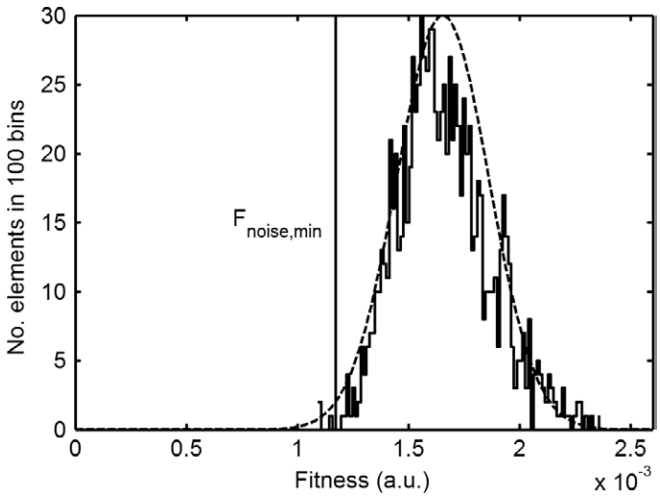


**Figure 4.** Histogram of 1000 fitness values in 100 bins (black solid curve) when the fitness is calculated between a noiseless and noisy XRR curves. The gaussian function approximating the histogram is presented in a black dashed line. The black vertical line shows the determined position of $F_{\text{noise,min}}$ in the fitness axis.

(v) Increase the index $i$ by one.
(vi) If $i < n$ where $n \gg 100$, then go to step (ii).

To maximize the right side of inequality (11), one can define that $\text{Min}[F(x_{\text{fit}}, \overline{x}_{\text{fit}})] = F_{\text{noise,min}} = \text{Min}(g)$. This definition, however, produces volatile values since $\text{Min}(g)$ may vary between simulations. A statistically more robust estimate for $F_{\text{noise,min}}$ can be obtained using the equation

$$F_{\text{noise,min}} = \mu_g + \sigma_g \Phi_g^{-1}(\alpha), \quad (13)$$

where $\Phi_g^{-1}(\alpha)$ is an inverse Gaussian cumulative distribution function and $\mu_g$ and $\sigma_g$ are the mean and the standard deviation of $g$, respectively. In this paper $\alpha = 0.01$ was used in equation (13) since it produced values nearly equal to $\text{Min}(g)$. Figure 4 illustrates the histogram of $g$, the Gaussian function approximating $g$ and $F_{\text{noise,min}}$ determined from equation (13).

Note that the right side of inequality (11) is positive but the left side equals zero with the first guess $\overline{x}_{\text{opt}} = \overline{x}_{\text{fit}} = \overline{x}(p)$. In order to obtain the maximum of right side, $\overline{x}(p)$ is optimized as a function of the XRR parameter set $p$ using the equation:

$$F(\overline{x}, \overline{x}_{\text{fit}}) = F(x_{\text{meas}}, \overline{x}_{\text{fit}}) - F_{\text{noise,min}}, \quad (14)$$

where the function $\overline{x} = \text{calcXRRcurve}(p)$. To find a $p$, the following procedure must be carried out find an upper limit for the error:

(i) Set a search direction $d$.
(ii) Set $p = p_{\text{fit}}$.
(iii) Compute $F_{\text{noise,min}}$ at $p$.
(iv) Optimize $p$ in the direction of $d$ until equation (14) is satisfied.

One can note that $F_{\text{noise,min}}$ is a function of $p$, strictly speaking. However, preliminary studies showed that $p$ has no observable effect on $F_{\text{noise,min}}$ and therefore the dependence was not taken into account.

*3.2.2. Implementation of the method.* The error limit determination method was implemented here in circular coordinates in two dimensions due to computational reasons. The parameter set $p = [t, \rho, \sigma]$ where $t$ is the layer thickness, $\rho$ is the mass density and $\sigma$ is the surface roughness of the layer to be optimized. The implementation was realized as follows:

(i) Select the plane, ($\rho$,$\sigma$) for instance, where to define error limits.
(ii) Compute vector $g$ with the parameters $p_{\text{fit}}$ and determine $F_{\text{noise,min}} = F_{\text{noise,min}}(p_{\text{fit}}, \alpha)$.
(iii) Define a search direction $\text{d}(\theta)$, where $\theta \in [0^\circ, 360^\circ]$. In ($\rho$,$\sigma$) plane

$$\text{d}(\theta) = p \begin{bmatrix} 0 & 0 & 0 \\ 0 & \cos(\theta) & 0 \\ 0 & 0 & \sin(\theta) \end{bmatrix}. \quad (15)$$

(iv) For the given $\theta$,
  (a) perform $r_{\max} = \text{Minimize}_r[F(x_{\text{meas}}, \overline{x}_{\text{fit}}) - F_{\text{noise,min}} - F(\overline{x}, \overline{x}_{\text{fit}})]$ with $p = p_{\text{fit}} + rd$,
  (b) save $p_{\text{contour}}(\theta) = p_{\text{fit}} + r_{\max} d$.
(v) Draw a closed contour $p_{\text{contour}}$ in the selected plane.

## 4. Results and discussion

### *4.1. Validity of the results*

In order to test the validity of the results given by the implementation, the error was determined for the layer structure shown in figure 1. In the task, the simulations were carried out using 1000 fitness value computations for the vector $g$ and 512 search directions between 0° and 360°. The 'measured' fitness was set to $F(x_{\text{meas}}, \overline{x}_{\text{fit}}) = F_{\text{noise,max}}$ where $F_{\text{noise,max}} = \mu_g + \sigma_g \Phi_g^{-1}(\alpha = 0.99)$. This selection was due to the observation that it is difficult to get better fitness than $F_{\text{noise,max}}$ to a real world measurement.

It is known by the skilled experimentalists that the error in the determined parameters follow approximately $C_{\max}^{-1/2}$ dependence if the averaging time was kept constant for every point. The linear dependence holds for the $\chi^2$ measure as shown in figure 5. After switching $\chi^2$ to $F$, the results exhibits more nonlinear behaviour as shown in figure 6 but increases as a function of $C_{\max}^{-1/2}$ which is required. Thus, one can conclude that the implementation used in the error limit determination gives expected results thus indicating valid realization of the method.

### *4.2. Results*

The errors were computed for XRR parameters determined from the real-world atomic-layer-deposited AlO layer. The maximum numbers of photon counts was $C_{\max} = 4.4 \times 10^7$. Figure 7 shows the determined error in the ($\rho$, $\sigma$) plane. The errors are $\Delta\rho = 0.105\,37\,\text{g cm}^{-3}$ and $\Delta\sigma = 0.063\,634\,\text{nm}$. Note that the area of the contour is nonparaxial which means that monovariate techniques may fail in the error limit determination.

Figure 8 shows a contour restricting the error region in the ($t$, $\sigma$) plane which is paraxial. The errors in this case are

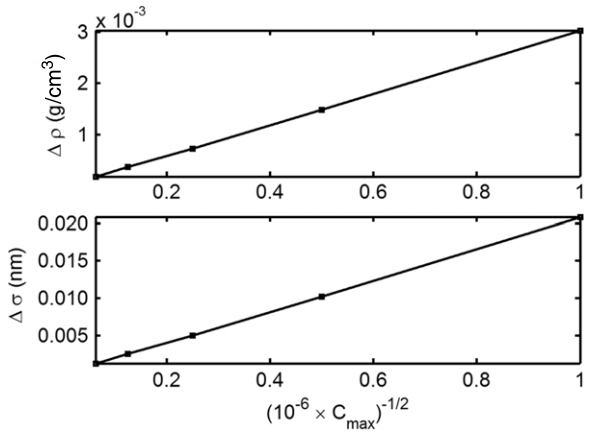


**Figure 5.** Error determined by using the $\chi^2$ measure as a function of $C_{\max}$ for the mass density $\rho$ and the surface roughness $\sigma$.

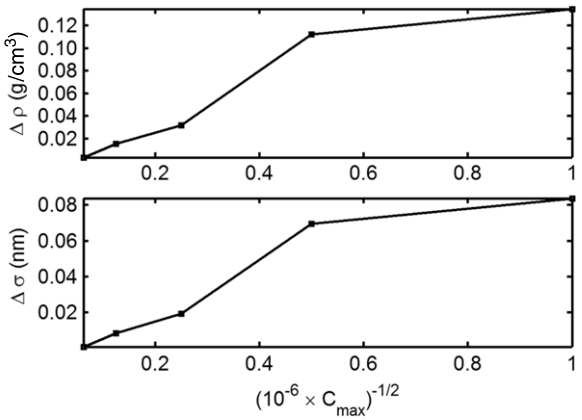


**Figure 6.** Error determined by using the $F$ measure as a function of $C_{\max}$ for the mass density $\rho$ and the surface roughness $\sigma$.

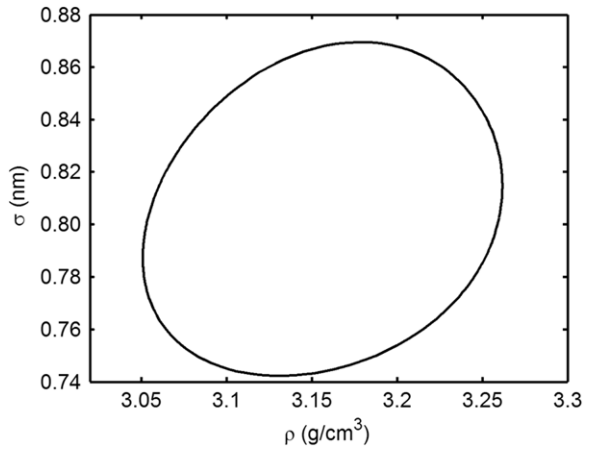


**Figure 7.** Error limits for the mass density $\rho$ and surface roughness $\sigma$. The black line encloses the area where other possible solutions can lie.

$\Delta t = 0.122\,87\,\text{nm}$ and $\Delta\sigma = 0.061\,591\,\text{nm}$. Figure 9 shows the error limits in the ($t$, $\rho$) plane and here $\Delta t = 0.125\,66\,\text{nm}$ and $\Delta\rho = 0.1051\,\text{g cm}^{-3}$. Note that the principal axes of the ellipse are not paraxial. This clearly indicates that the mass density and the thickness are slightly correlating, i.e. the thickness is not a fully independent variable in the analysis as thought in general.

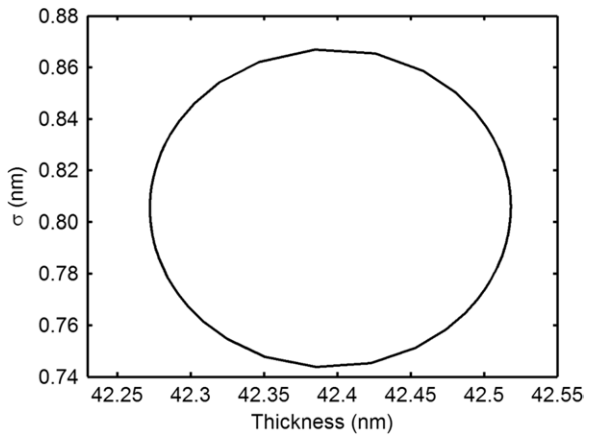


**Figure 8.** Error limits for the thickness $t$ and surface roughness $\sigma$. The black line encloses the area where other possible solutions can lie.

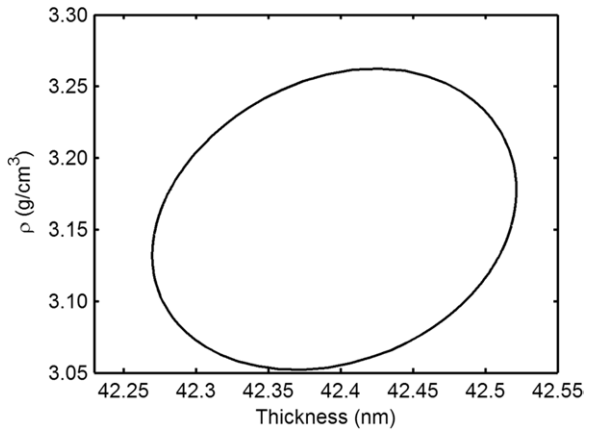


**Figure 9.** Error limits for the thickness $t$ and the mass density $\rho$. The black line encloses the area where other possible solutions can lie.

XRR curve fitting was performed using the $\chi^2$ measure. $\chi^2$ is very sensitive to the mass density and it can be considered to be more accurate than the mass density determined using $F$. Table 1 shows the determined XRR parameters for both measures, the differences in the parameters and the computed errors for both measures. As seen from figure 2, $\chi^2$ cannot match the right tail of the curve and therefore the difference in the thicknesses is considerably higher than the calculated error using $F$. On the other hand, the error determined using $\chi^2$ is considerably large which indicates the failure of the fit in the thickness determination. The error in the mass density value determined using $\chi^2$ is pretty small indicating the aforementioned accuracy of the measure with the mass density. It is interesting to note that the errors together agree with the difference in the determined mass density parameters. In the case of the surface roughness, the error determined using $\chi^2$ is clearly higher than in the case of $F$ indicating again the failure of the fit done using $\chi^2$. It is worth recalling here that the determined errors using $\chi^2$ is done here for comparison purposes but strictly speaking, the results are valid only for $F$. This is due to the lack of error derivation for $\chi^2$ and therefore inequality (11) does not necessarily hold for $\chi^2$. However, some inaccuracy may exist in the results obtained by using $F$.

**Table 1.** Determined XRR parameters with $F$ (denoted as $^\dagger$) and $\chi^2$ (denoted as $^\ddagger$). The XRR curves based on these parameters are presented in figures 2 and 3 for $\chi^2$ and $F$, respectively. Differences are the parameter deviations in the principal $Al_2O_3$ layers. Here $^*$ denotes a parameter which was fixed during the fitting procedure. Error$^\dagger$ is calculated using the $F$ measure. The parameters are here presented with three decimal accuracy for the thickness $t$, the AlO mass density $\rho$ and the surface roughness $\sigma$.

| Layer | $t$ (nm) | $\rho$ (g cm$^{-3}$) | $\sigma$ (nm) |
|---|---|---|---|
| $Al_2O_3{}^\dagger$ | 42.393 | 3.152 | 0.804 |
| Subtrate$^\dagger$ | $\infty$ | 2.33$^*$ | 0.000 |
| $Al_2O_3{}^\ddagger$ | 42.152 | 3.044 | 0.850 |
| Subtrate$^\ddagger$ | $\infty$ | 2.330$^*$ | 0.000 |
| | $\Delta t$ (nm) | $\Delta\rho$ (g cm$^{-3}$) | $\Delta\sigma$ (nm) |
| Difference | 0.242 | 0.108 | 0.046 |
| Error$^\dagger$ | 0.123 | 0.105 | 0.064 |
| Error$^\ddagger$ | 0.653 | 0.033 | 0.215 |

Recall here that the error analysis is based on inequality (11) where a sufficiently large number of datapoints and signal-to-noise ratio were assumed. Therefore, the inaccuracy of results given by the novel method increases if these conditions are not met.

### *4.3. Discussion*

Here the presented novel technique is discussed in the light of the classical Hessian based sensitivity method. The Hessian method is based on the examination of the Hessian matrix which shows the local sensitivity of the fitted parameters. Hessian is here defined as

$$H(F) = \left.\frac{\partial^2 F[x_{\text{meas}}, \overline{x}(p_i, p_j)]}{\partial p_i \partial p_j}\right|_{p_{\text{fit}}}, \tag{16}$$

where $\partial p_i$ and $\partial p_j$ are the partial derivatives with respect to the parameters $i$ and $j$ and $p_{\text{fit}}$ is the parameter set of the fitted curve. In the examined case the numerical Hessian

$$H(F) = \begin{bmatrix} 1515.3 & -23.2 & 6.0 \\ -23.2 & 9.0 & -1.2 \\ 6.0 & -1.2 & 2.4 \end{bmatrix}, \tag{17}$$

where the partial derivatives are taken with respect to the layer thickness $t$, the mass density $\rho$ and the surface roughness $\sigma$. The eigenvectors of Hessian are

$$V(F) = 10^{-3} \begin{bmatrix} 1.3 & -15.9 & 999.9 \\ -172.9 & -984.8 & -15.4 \\ -984.9 & 172.9 & 4.0 \end{bmatrix}, \tag{18}$$

where the vectors $[t, \rho, \sigma]^{\text{T}}$ are in columns. The eigenvalues are 2.2, 8.8 and 1515.7 for the column vectors, respectively. The direction of an eigenvector with a small eigenvalue means little change in $F$ and hence larger uncertainty or error for this linear combination of parameters. Therefore the directions $[t, \rho, \sigma]^{\text{T}} = [1.3, -172.9, -984.9]^{\text{T}}$ and $[-15.9, -984.8, 172.9]^{\text{T}}$ give the largest uncertainty which

means that the error is greatest due to the crosscorrelation between the mass density and the surface roughness. The direction of the principal axes of the ellipse in figure 7 indicates crosscorrelation between the mass density and the surface roughness, respectively. Thus the results of these two methods agree qualitatively in this case. If the uncertainty is studied using the Hessian method in the direction of $[999.9, -15.4, 4.0]^{\mathrm{T}}$, the mass density interacts with the thickness more strongly than the surface roughness with the thickness, as also indicated by the novel results shown in figures 8 and 9. Thus one can conclude that the results of the novel method agree qualitatively with the results of the Hessian approach.

The quantitative comparison of the Hessian based error analysis and the novel method is difficult due to the differences between the techniques. The differences are:

- The Hessian method uses the *measurement and the fit* in the analysis whereas the novel method studies the difference between *the fit and an unknown optimal fit*.
- The Hessian method studies the *sensitivity* of the fitted parameters whereas the novel method studies *noise* affecting fitness values.
- The Hessian method *preserves* the information of the measurement during the analysis but the novel method *loses* this information.
- If the Hessian method is used for the error limit determination,
  - it *assumes* that the fitted solution gives the minimum fitness. This is in fact an incorrect assumption since the fitness of the fitted solution does not even fall into the distribution shown in figure 4. The novel method does not make such an assumption.
  - the suitable value for $\epsilon$ is difficult to define in the case of the nonideal fit. The novel method takes into account a nonideal fit and does not require difficult/complex definitions for any parameter.

Summarizing the differences between the methods, the Hessian method answers the question 'How an error in one parameter affects the rest of parameters?' but the novel method answers the question 'If the search algorithm has not minimized the fitness value, where or how far the optimal solution can be ultimately?'. Since these methods have different perspectives, the methods are competing each other thus providing valuable information on the uncertainties of the problem.

## 5. Conclusions

The novel method taking into account the influence of noise and nonideal fit in the error limit determination in XRR analysis was presented. The method applies Parratt's formalism and the fitting based on least squares of logarithms. This logarithmic approach was selected since it gave good visual correspondence between the measurement and the fit. The fitness measuring the correspondence between the measurement and the fit was formally divided into two parts where the first part represents the fitness between the fit and an unknown optimal fit and the second part is the fitness of noise. The fitness of noise was approximated to be as minimal as possible with a simple separate simulation thus increasing the fitness of a nonideal fit. In this minimization task, the statistical significance level $\alpha = 0.01$ was used as the limit which separates 99% of the simulations to have a greater fitness value than the reference value, i.e. the fitness of noise. The error limit determination was performed by increasing the distance between the fit and an unknown optimal fit in the parameter space until the fitness between these corresponding curves met the original fitness minus the fitness of noise.

One simple and computationally efficient implementation for the practical error limit determination was presented in this work but other approaches can be used as well. Using the presented implementation, initial simulations were made to test the validity of results. The error limits were halved with quadrupled intensity which indicated the correct behaviour of the implementation. Finally, the error limits were determined for a real-world atomic-layer-deposited aluminium oxide layer on silicon as an application. The determined parameters were $42.4 \pm 0.12$ nm (0.3%), $3.15 \pm 0.11$ g cm$^{-3}$ (3.5%) and $0.80 \pm 0.06$ nm (7.5%) for the thickness, the mass density and the surface roughness, respectively.

The error limits given by the novel method are upper limits in the case of a sufficient number of sampling points and photon counts. In practice, the determined error is an asymptotic estimate and requires fine tuning as future work. However, the significance of the approach is related to its capability to take into account a systematic error caused by a nonideal fit which may be difficult with classical methods. The novel method requires the definition of two parameters, the statistical significance level $\alpha$ and the number of simulations for the distribution of the fitness of noise. The number of simulations affects the accuracy of the results and should, in principle, be as large as possible. The significance level is a user-defined parameter and it can be selected to meet the required confidence level. Since the presented novel method does not contain arbitrary parameters, it gives objectively an estimate for the upper limit of the error when the convergence of a search algorithm to the optimal fit is not guaranteed.

## Acknowledgments

The authors thank 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 thanked for providing IT services.

## Appendix

The fitness between the measurement and the fit can be written as

$$\begin{aligned} F(x_{\mathrm{meas}}, \overline{x}_{\mathrm{fit}}) &= N^{-1} \parallel e_{\mathrm{min}} + e_{\mathrm{res}} \parallel_2^2 \\ &= N^{-1} \parallel e_{\mathrm{min}} \parallel_2^2 + N^{-1} \parallel e_{\mathrm{res}} \parallel_2^2 \\ &\quad + 2N^{-1} e_{\mathrm{min}} \cdot e_{\mathrm{res}} \\ &\geqslant N^{-1} \parallel e_{\mathrm{min}} \parallel_2^2 + N^{-1} \parallel e_{\mathrm{res}} \parallel_2^2 \\ &\quad - 2|N^{-1} e_{\mathrm{min}} \cdot e_{\mathrm{res}}| \end{aligned} \tag{A.1}$$

where $e_{\min} = \log x_{\text{meas}} - \log \overline{x}_{\text{opt}}$ and $e_{\text{res}} = \log \overline{x}_{\text{opt}} - \log \overline{x}_{\text{fit}}$. Hence

$$N^{-1} \parallel e_{\text{res}} \parallel_2^2 \leqslant N^{-1}(\parallel e_{\min} + e_{\text{res}} \parallel_2^2 - \parallel e_{\min} \parallel_2^2) + 2|N^{-1} e_{\min} \cdot e_{\text{res}}|. \tag{A.2}$$

Now

$$\begin{aligned} e_{i,\min} &= \log x_{i,\text{meas}} - \log \overline{x}_{i,\text{opt}} \\ &= \log \frac{x_{i,\text{meas}}}{\overline{x}_{i,\text{opt}}} \\ &\approx \log \frac{C_{\max}^{-1}\text{Poisson}(C_{\max}\overline{x}_{i,\text{opt}})}{\overline{x}_{i,\text{opt}}}. \end{aligned} \tag{A.3}$$

Note here that $C_{\max}^{-1} \gtrsim \overline{x}_{i,\text{opt}} \leqslant 1$ and $C_{\max}^{-1} \geqslant \overline{x}_{i,\text{opt}} \lesssim 1$ for all $i$ and therefore the variance of $e_{i,\min}$ is bounded. Let us denote

$$e_{\min} \cdot e_{\text{res}} = \sum_{i=1}^{N} e_{i,\min} e_{i,\text{res}} = \sum_{i=1}^{N} E_i \tag{A.4}$$

and $\sigma_i^2 = \mathbb{E}[E_i^2] > 0$ where $\mathbb{E}$ is an expectation operator. Since the XRR data is preprocessed, then for all $i$

$$\sigma_i^2/\sigma^2 < \epsilon \tag{A.5}$$

is satisfied for small $\epsilon > 0$, where

$$\sigma^2 = \sum_{k=1}^{N} \sigma_k^2. \tag{A.6}$$

Now the Lindeberg condition is satisfied for the central limit theorem which states that the sum distribution is approximately Gaussian, i.e.

$$\begin{aligned} &\text{Prob}\left(\frac{\sum_{i=1}^{N}(E_i - \mathbb{E}[E_i])}{\sigma} \leqslant a\right) \\ &\to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \exp\left(\frac{-u^2}{2}\right) \text{u}, \qquad N \to \infty, \end{aligned} \tag{A.7}$$

where the left side denotes the probability of normalized sum distribution less than $a$ and the right side is the standard cumulative Gaussian distribution. Thus

$$N^{-1} \sum_{i=1}^{N} E_i \to N^{-1} \sum_{i=1}^{N} \mathbb{E}[E_i] + N^{-1}\sigma X, \tag{A.8}$$

where $X$ is a random number from standard Gaussian distribution with zero mean and unit variance. Here $N^{-1}\sigma X = c^{1/2}N^{-1/2}X \to 0$ for large $N$ where $c$ is a bounded mean variance. Therefore

$$N^{-1} \sum_{i=1}^{N} E_i = N^{-1} \sum_{i=1}^{N} \mathbb{E}[E_i]. \tag{A.9}$$

No nice analytical formula exists for $\mathbb{E}[E_i]$. However, for sufficiently large $C_{\max}\overline{x}_{i,\text{opt}}$

$$e_{i,\min} \approx \log \frac{C_{\max}^{-1}\text{Poisson}(C_{\max}\overline{x}_{i,\text{opt}})}{\overline{x}_{i,\text{opt}}} \tag{A.10}$$

$$\approx \log \frac{C_{\max}^{-1}[C_{\max}\overline{x}_{i,\text{opt}} + \sqrt{C_{\max}\overline{x}_{i,\text{opt}}}X]}{\overline{x}_{i,\text{opt}}} \tag{A.11}$$

$$= \log\left[1 + \frac{1}{\sqrt{C_{\max}\overline{x}_{i,\text{opt}}}}X\right], \tag{A.12}$$

where Poisson's distribution is approximated using the normal distribution. If one considers the case when $C_{\max}\overline{x}_{i,\text{opt}} \geqslant 100$ and $X \in [-1, 1]$ with 68% probability, then $\ln(1-t) \approx -t$ for small $|t|$, i.e.

$$\mathbb{E}[E_i] \approx \mathbb{E}\left[e_{i,\text{res}} \frac{1}{\ln(10)\sqrt{C_{\max}\overline{x}_{i,\text{opt}}}}X\right] \tag{A.13}$$

$$= e_{i,\text{res}} \frac{1}{\ln(10)\sqrt{C_{\max}\overline{x}_{i,\text{opt}}}}\mathbb{E}[X] = 0, \tag{A.14}$$

where $e_{i,\text{res}} \approx 0$ is the slowly varying parameter as a function of $i$, i.e. constant and

$$\begin{aligned} N^{-1} \parallel e_{\text{res}} \parallel_2^2 &\leqslant N^{-1}(\parallel e_{\min} + e_{\text{res}} \parallel_2^2 - \parallel e_{\min} \parallel_2^2) \\ &\quad + 2|N^{-1} e_{\min} \cdot e_{\text{res}}|, \\ &\to N^{-1}(\parallel e_{\min} + e_{\text{res}} \parallel_2^2 - \parallel e_{\min} \parallel_2^2), \end{aligned} \tag{A.15}$$

for sufficiently large $N$ and $C_{\max}\overline{x}_{i,\text{opt}}$. Note that these conditions are met when the number of sampling points and an averaging time in a measurement are simultaneously increased. If the conditions are not met, then the upper bound for the error can be fine tuned by taking into account the mean of expectation values. However, we leave this case to be studied in future work.

## 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] Filippino A 1995 *J. Phys.: Condens. Matter* **7** 9343–56
[11] Tiilikainen J, Bosund V, Mattila M, Hakkarainen T, Sormunen J and Lipsanen H 2007 *J. Phys. D: Appl. Phys.* **40** 4259–63
[12] Tiilikainen J, Tilli J-M, Bosund V, Mattila M, Hakkarainen T, Sormunen J and Lipsanen H 2007 *J. Phys. D: Appl. Phys.* **40** 7497–501
[13] Puurunen R L 2005 *J. Appl. Phys.* **97** 121301