Thin Solid Films 595 (2015) 113–117

Contents lists available at ScienceDirect

Thin Solid Films

journal homepage: www.elsevier.com/locate/tsf

A practitioner's approach to evaluation strategy for ellipsometric measurements of multilayered and multiparametric thin-film structures

D.V. Likhachev

GLOBALFOUNDRIES Dresden Module One LLC & Co. KG, Wilschdorfer Landstr. 101, D-01109 Dresden, Germany

ARTICLE INFO

Article history:
Received 29 September 2015
Accepted 30 October 2015
Available online 4 November 2015

Keywords:
Data analysis
Model selection
Parameterization
Information criteria
Spectroscopic ellipsometry
Optical characterization
Multilayered structures
Optical metrology

ABSTRACT

Ellipsometry as an indirect optical measurement method requires the use of optical modelling which include model parameterization. In practice, there are many ways to select a model and its parameters to fit the experimental data. Very often this fact leads to ad hoc decisions, i.e., based on experience or subjective opinion, instead use of some systematic approaches which provide predictive capability. In this paper we use the Akaike and Bayesian information criteria to perform optical model selection and its best parameterization to fit a particular set of ellipsometric data. We demonstrate that this approach accompanied by *post hoc* study of the inter-parameter correlations can significantly enhance optical modelling, in particularly, the process of model selection and data interpretation and improve the characterization of multilayered thin-film structures.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

As a search of the literature points out, Dr. Jellison's 1993 paper on data analysis in spectroscopic ellipsometry (SE) [1] is one of the most cited research articles in the area of SE data interpretation (see also Refs. [2–4]). There he formulated in consecutive order three-phase procedure of fitting a model to experimental ellipsometric data:

1. Construct a model by defining the number of layers and their types — homogeneous or inhomogeneous (graded), isotropic or anisotropic, with or without interface layers.
2. Select the optical functions for each layer (for instance, as tabulated data) or parameterize them if the material optical properties are unknown a priori and need to be determined during characterization.
3. Fit the model, in which a few parameters are allowed to vary, to the experimental data using appropriate multi-parameter non-linear optimization algorithm and minimizing so-called *merit function* (or *error function*), i.e., the function which determines the quality of fit.

The need for advanced fitting procedures has been long recognized in the applied-math community. The situation has been well reflected by Press et al. [5, pp. 498–499]: "…it is not uncommon in fitting data to discover that the merit function is not unimodal, with a single minimum. In some cases, we may be interested in global rather than local questions. Not, "how good is this fit?" but rather, "how sure am I that there is not a very much better fit in some corner of parameter space?"… The important message we want to deliver is that fitting of parameters is not the end-all of model parameter estimation. To be genuinely useful, a fitting procedure should provide (i) parameters, (ii) error estimates on the parameters, and (iii) a statistical measure of goodness-of-fit. When the third item suggests that the model is an unlikely match to the data, then items (i) and (ii) are probably worthless. Unfortunately, many practitioners of parameter estimation never proceed beyond item (i). They deem a fit acceptable if a graph of data and model "looks good." This approach is known as *chi-by-eye*. Luckily, its practitioners get what they deserve." Yet, we believe that these expressive words are still relevant to the community of ellipsometry users, — especially, "typical" industrial ellipsometry users as defined by Harland Tompkins in Ref. [6]; hence, there is a hard need for well-developed strategies in ellipsometric measurement evaluations.

Various advanced optimization algorithms for evaluation of the ellipsometric measurements have been already discussed in great details by Polgár et al. [7–10] (see also, for instance, Ref. [11, pp. 196–203]). For the purpose of modelling or parameterization many analytical physics-based and Kramers–Kronig consistent expressions (models) have been developed which describe various types of materials — amorphous and crystalline semiconductors and dielectrics, metals, organic films, optical metamaterials, etc. [12,13]. Parameterization should be accompanied by sensitivity analyses [14–16]. Sensitivity analysis determines the relative importance of the parameters and helps to optimize a range of variations for each sensitive parameter since the efficiency of all optimization algorithms can be greatly improved if the parameter search space has reasonable bounds. Unfortunately, in general, it seems that there is a lack of

E-mail address: dmitriy.likhachev@globalfoundries.com.

http://dx.doi.org/10.1016/j.tsf.2015.10.078
0040-6090/© 2015 Elsevier B.V. All rights reserved.

systematic estimation and validation *by all available means* of which model and its parameterization are more adequate to describe the measured data. The current paradigm has been well remarked by Herzinger et al. [17]: "When performing a model dependent data analysis, simple models are preferable to complex ones if the fit quality is the same… The difficulty is in objectively evaluating the quality of the fit as each model complication is added to see if the fit really improved. Of course, if the fit does not get better with increasing complexity, that does not mean the complex model is necessarily wrong, but it does mean that one lacks sensitivity to allow a distinction and some other criteria must be employed." As a heuristic approach, if too many unknown parameters are selected to vary, an overfitting of the spectral data may occur easily, i.e., small variations in the experimental data may strongly influence the parameter estimates and predictions from the model become questionable. We should always bear in mind a famous jest by von Neumann: "…with four parameters I can fit an elephant, and with five I can make him wiggle his trunk." (attributed to John von Neumann by Enrico Fermi, as quoted by Freeman Dyson in Ref. [18]). As a matter of fact, Mayer et al. [19] confirmed von Neumann's assertion by reconstructing an elephantine shape with four complex numbers and even making the trunk wiggle using the real part of the fifth parameter. At the opposite extreme, an insufficient number of variables or an inflexible/inadequate model (which is less adapted to more complicated situations than a flexible model) will produce an underfitting which will also result in unreliable predictions. Also, typically, there are a few candidate models which are physically valid and can be used to fit the measured data. Final selection from those competing models is usually based on a value of χ^2, the biased or unbiased estimator of the goodness of fit [20], as well as the estimations of the cross-correlation coefficients and confidence limits of the best-fitting parameters. However, as many practitioners already gained from their experience, it is possible to devise different models, with different or even equal number of adjustable parameters (i.e., equal complexity), which will result in very similar quality of fit. *In spite of that, those models may produce different results when applied to a given set of measured data and the model's predictions become less accurate.* Therefore, a choice of the model as well as the number of parameters to adequately represent the data is an important problem which still doesn't have perfect solutions.

To address this problem a variety of statistical criteria for model selection and parameter estimation have been derived and applied in various disciplines, including cosmology, quantum information theory, life sciences, social sciences, medicine, economics, and numerous industrial applications. One such long-standing and widespread measure, called Akaike information criterion (AIC), has been established in the early 1970s [21,22] and it is founded on the information-theoretic approach [23–25]. Another commonly used statistical criterion is Bayesian information criterion (BIC) [26] introduced by Schwarz [27] (also known as Schwarz or SIC criterion). Both of these methods allow finding a compromise between under- and overfitting of experimental data. Before applying any of these criteria, a set of possible models with different numbers of parameters needs to be created. In fact, the AIC and BIC penalize for the addition of parameters, i.e., increasing model complexity, and selects a model which achieves the best fit with minimal number of parameters. Both AIC and BIC are relatively easy to use since they are described by simple formulas and require only a value of the residual difference between measured and calculated data which is one of the outputs of a given model and reflects the overall quality of fit of the model. An important advantage of using AIC and BIC model-selection criteria is that they allow to make inferences based not only on model's goodness of fit but also on parametric complexity of the model. Thus, the statistical criteria approach provides an objective way to yield relative rather than absolute ranking of candidate models and identify the most appropriate one, — the task that has always been ambiguous for most of the practitioners. But we also should keep in mind that these criteria do not identify the "true" or "final" model, i.e., if a set of the models under test contains only inferior models, the AIC or BIC will select the best one out of the poor models [23, p. 62]. As stated by George Box, "Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful." [28]. In other words, any model that we introduce is considered, at best, only as an approximation of reality. However, purely statistical model-selection scenario should be applied with some caution since not only statistics but underlying physics or technological aspects need to be taken into account in ellipsometric measurement evaluations.

Despite of wide range of fields where the AIC and BIC were applied for model comparison, those methods so far have not been utilized in ellipsometry. In this paper we use the Akaike and Bayesian information criteria to perform optical model evaluation and its best parameterization. We show by example the use and interpretation of the results of AIC and BIC model-selection procedures with a particular set of ellipsometric data. We also demonstrate that the model-selection approaches should be supplemented also by some kind of cross-validation methods, such as inter-parameter correlations study.

2. The information criteria

Let us assume that the experimental data consist of n data points y_i taken at different values of independent variable x, i.e., we have n pairs of (x_i, y_i). Now suppose that we can fit those measured data with m different functional relations (candidate models) $f_j(x_i|\theta)$, $j = 1, \ldots, m$, where θ is the vector of parameters in the approximating model. If the residuals are normally distributed with zero mean value and experimental error σ, the quality of fit is quantified by the following measure.

$$\chi^2 = \frac{1}{2}\sum_{i=1}^{n}\frac{\left(y_i - f_j(x_i|\theta)\right)^2}{\sigma_i^2} = -2\ln\mathcal{L} + C, \quad (1)$$

where $\mathcal{L}$ is the likelihood function and the constant C depends only on used set of experimental data points and not on the model.

The AIC and BIC are defined as

$$\mathrm{AIC}_j = -2\ln\mathcal{L}_{\max,j} + 2p, \quad (2)$$

$$\mathrm{BIC}_j = -2\ln\mathcal{L}_{\max,j} + p\ln n, \quad (3)$$

where $\mathcal{L}_{\max,j}$ is the maximum likelihood of an estimated model j from a set of candidate models, M_j ($j = 1,\ldots, m$), p is the total number of model parameters, and n is the number of data points used in the fit [21,22,27]. Despite apparent similarity in the formulas for AIC and BIC, they were derived within different theoretical frameworks (for brevity the derivation of AIC and BIC had to be omitted and we refer the interested reader for technical details on the information-theoretic and Bayesian approaches to Refs. [23–25]). Though, both the AIC and BIC measures have two terms. The first one is a measure of the model lack of fit and can be reduced by increasing number of parameters in the model. At the same time, the second term penalizes for the additional parameters in the model and increases with increasing number of parameters. The model which gives the minimal AIC or BIC value (score) is considered as the best fitting model since it minimizes the difference between the candidate model and the measured data by minimal number of parameters. Clearly, the BIC penalizes for the addition of new parameters more strictly than the AIC due to the presence of ($p \ln n$) penalty weight term and, therefore, it tends to select simpler models, i.e., the models with a smaller number of parameters (since $\ln n > 2$ for any practicable data sets and if only $n = 7$ or 8, then BIC ≈ AIC). Usually, these both criteria have shown good agreement on the ranking of candidate models which is fairly surprising since these criteria represent very different approaches [29]. If, in some instances, the results of model selection from two criteria disagree, one can have a positive outlook on that fact as on suggested bounds for the range of acceptable models [30]. Therefore, for instance, AIC provides an upper bound to the number of

Download English Version:

https://daneshyari.com/en/article/1664260

Download Persian Version:

https://daneshyari.com/article/1664260

Daneshyari.com