International Journal of Applied Mathematics

Volume 28 No. 1 2015, 85-98
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v28i1.7

GLOBAL OPTIMIZATION IN INVERSE ELLIPSOMETRIC PROBLEM FOR THIN FILM CHARACTERIZATION

Lekbir Afraites

Laboratoire de Math. et Appl.
Fac. des Sci. et Techn.
University Sultan Moulay Slimane
P.O. Box 523, Beni-Mellal, MORROCO

Abstract: In the current work, we consider the inverse ellipsometric problem for thin film characterization which consists in determining the shape of a diffracting feature from an experimental ellipsometric data. The reformulation of the given nonlinear identification problem was considered as a parametric optimization problem using the Least Square objective function. The evaluation of the latter is often expensive as it implies the solution of the direct problem for each iteration. In this work, we propose a design procedure for global robust optimization using a probabilistic approach based upon Kriging method. Robustness is determined by the Kriging model to reduce the number of real functional calculations of Least Square objective function. The technical of the global optimization methods is adopted to determine the global robust optimum of a objective function considered.

AMS Subject Classification: 90C31, 90C59, 97N50, 90C26, 90C27
Key Words: inverse ellipsometric problem, Kriging method, least squares, global optimization

1. Introduction

The process control in microelectronics manufacturing requires real time monitoring techniques. Among the different metrology techniques, scatterometry, based on the analysis of the light diffracted by microscale patterns using for example an ellipsometer, is well suited. The problem of computing the signa-

Received: December 5, 2014

ture from a given structure shape, which is referred to as the direct problem, is dealt with using conventional Maxwell equations solvers, generally based on modal methods [10]. On the opposite, the inverse problem [11, 4], which allows the determination of the feature shape from an experimental signature is solved by parametric optimization problem using the Least Square objective function. This inverse problem is difficult to solve. On one hand, the problem is ill-posed, which requires for example the use of regularization methods. On the other hand, the use of traditional optimization methods brings us back to a local optimum and the quality of the result depends on the initial point. To solve this problem, one generally use [13, 14] precomputed library and look for the best matching signatures inside this library (hence the supposedly best parameters). Among the disadvantages of this method, the computing time of the direct problem is extremely expensive. In order to deal with the issue of local optimum and the dependence of the initial point, we propose an approach based on the Kriging interpolation method and its use as a technique for global optimization. The paper is organized as follows. In the second section, we present the principles of ellipsometric scatterometry and then describe the direct problem. In Section 3, we present the inverse problem which we will consider in the present work. In Section 4, we present the efficient global optimization (EGO) [6] algorithm. It sequentially samples results from an expensive calculation, does not require derivative information and uses an inexpensive surrogate itself obtained by Kriging method to search for a global optimum. In the final section, we present an application of the algorithm EGO on a synthetic example and the inverse problem of ellipsometry where in the first example, we determine the shape of a diffracting feature from the synthetic data and the second example we use experimental data delivered by the LTM-laboratory Grenoble, France.

2. Direct Problem: Ellipsometric Signature

Scatterometry is used as a generic term for several metrology methods. It may be described as a measurement technique allowing for a quantitative evaluation of the geometrical or material properties of an object through the analysis of the light scattering from the surface under test. Since no imaging optics is used, the surface and the shape have to be reconstructed from intensity and/or polarization data detected in the far field. In our case, we use the spectroscopic ellipsometry. The metrology device that measures the polarization change upon reflection by the sample is kept static whereas the incident wavelength is varying. As mentioned in the introduction, the direct problem is used to establish signatures from a given shape topography using a Maxwell solver. We use the Modal Method by Fourier Expansion to do that [10]. This method is well adapted for the rectangular topography of the samples used in the microelectronic manufacturing which are of primary interest for us. The direct problem computes theoretically the ellipsometric signal values Is and Ic (from Fresnel Equations), that will be denoted by Is_i^t and Ic_i^t where i represents the wavelength $(i \in N)$:

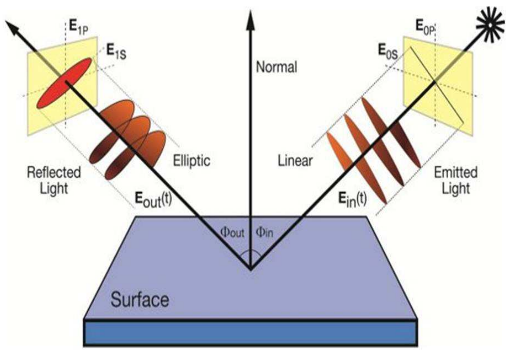

Figure 1: principe

$$Is^t = \sin 2\psi \sin \Delta, \quad Ic^t = \sin 2\psi \cos \Delta,$$
$$\text{for a 0 angle between the polarizer and the modulator,} \quad (1)$$

$$Is^t = \sin 2\psi, \quad \sin \Delta, \quad Ic^t = \cos 2\psi,$$
$$\text{for an angle } \tfrac{\pi}{4} \text{ between the polarizer and the modulator.} \quad (2)$$

In these two equations, ψ and Δ are computed as follows

$$\rho = \frac{r_p}{r_s} = \tan \psi \; e^{i\Delta}. \quad (3)$$

Equations (3) represents the formalism which allows to compute ρ, the ratio of complex reflectivities. r_p and r_s are the Fresnel reflection coefficients for the P and S polarization respectively. The ellipsometry measures the change in the amplitude ψ and phase Δ, between the p (parallel) and the s (perpendicular) components of polarized light upon reflection from a surface [3]. Once ψ and

Δ are known, one can theoretically compute Is^t and Ic^t using Equations (1) or (2), depending on the configuration of the ellipsometer.

3. Inverse Problem

The ellipsometer provides us with the measured data Is^m and Ic^m, for the interval $[1, ..., N]$ of wavelength (or energy-eV), where i represents the wavelength ($i \in N$) and m indicates that these are measured values. Our goal is to solve the inverse problem [10, 11] which allows determining of the feature shape from an experiment ellipsometric values Is^m and Ic^m. For this, We transform our inverse problem to an optimization problem where the unknowns are the feature shape. We fit our parameters to experimental data Is^m and Ic^m using parameter identification methods. Parameter identification utilizes optimization methods to find the best parameter values. We consider the least squares objective function can be written as a difference between the direct signal theoretically calculated and experimental values measured:

$$J(L) = \frac{1}{2} \sum_{i=1}^{i=N} \Big[(Is_i^t - Is_i^m)^2 + (Ic_i^t - Ic_i^m)^2 \Big], \tag{4}$$

where L is the set of parameters optimization (generally describing the pattern shape). The objective of this work is to find the global optimum for this objective function J using the response surface obtained by Kriging method. For more details on the Kriging see [12, 9]. In our study, we applied the Kriging technique for the reconstruction of the ellipsometric signatures [2] and we use the global optimization techniques [6] described in the next section. Recently, this technique was used to identify the parameters in a model of immune competition [1].

4. Kriging Method and Global Optimization Technique

4.1. Kriging Method

The Kriging method is exact interpolation method [12, 9], developed by Matheron and Krige [5], is based on the theory of regionalized variables. It is a stochastic interpolation, which has proved to be reliable when approximating deterministic behaviours [7]. Indeed, it attempts to obtain statistically the optimal prediction, i.e. to provide the best linear unbiased estimator. The basic premise of the Kriging interpolation method is that every unknown point can be estimated by the weighted sum of the known points. The method also provides a mechanism for estimating the interpolation error for any approximated point. Thus, the use of this interpolation method allows to enrich the number of usable data together with controlling the error of the added approximated values [2].

More precisely, given the previously computed values of the objective function $f(x_1), ..., f(x_n)$ at the data points $x_1, ..., x_n$, Estimating the value of variable x in a not sampled site by a linear combination of specific data

$$\tilde{f}(x) = \sum_{i=1}^{n} \lambda_i f(x_i),$$

where λ_i depend on the distance of the test point x from observed points. The Kriging technique is essentially a method of interpolation between known points that provides a mean prediction $\tilde{f}(x)$, as well as a measure of the variability of the prediction $s(x)$ (the standard error MSE).

4.2. Global Optimization Technique

This section is inspired by the work of Donald Jones et al. [6]. The idea is based on the optimization of the response surface constructed by a Kriging model. The simplest possible way would be to fit a surface to the response surface then to find the minimum of this surface. However, if we proceed that way, we can easily be lead to a local minimum, and we have no specific information on the uncertain areas of the approximated response surface. To overcome this issue, we must put some emphasis on sampling the surface where we are uncertain, this is inherently measured by the standard error of the predictor. To combine the search for local and global minimum and we take into account the uncertainties of the Kriging surfaces, we use a criterion based of the balance between local and global search. This criterion is knows as *expected improvement*. It was introduced in the literature in 1978 in [8]. The expected improvement criterion is computed as follows. Let $f_{min} = \min(f(x_1), ..., f(x_n))$ be the current best function value. The *improvement function* I at the point x is defined as follows:

$$I(x) = \max\Big(f_{min} - F(x), 0\Big), \tag{5}$$

where $F(x)$ is Normal $(\tilde{f}, s^2)$, i,e, F is a random variable with the mean and standard deviation given by the kriging predictor $\tilde{f}$ and its standard error s.

The expected improvement, defined as the expectation of the improvement, is given by [6]:

$$E[I(x)] = \Big(f_{min} - \tilde{f}(x)\Big)\phi\Big(\frac{f_{min} - \tilde{f}(x)}{s(x)}\Big) + s(x)\Phi\Big(\frac{f_{min} - \tilde{f}(x)}{s(x)}\Big), \tag{6}$$

where ϕ is the standard normal cumulative density function, and Φ is the standard normal probability density function.

This technique of global optimization consist of finding a good balance between minimum search and global exploration. Minimum search will result in a good approximation of the global minimum if the region of the global minimum is correctly represented by the response surface, and if this surface does not introduce minima lower than the global minimum. In order to guarantee a trustworthy representation, a more or less uniform coverage of the whole domain of interest should be generated.

4.3. Algorithm

The point where the value of the expected improvement is maximum gives the best point to evaluate the objective function. The expected improvement is constructed to search for both local and global minima [6]. The surrogate model is then updated to include the newest sampled point, and the operation is repeated until the sampling point does not change and the global minimum of the objective function has been found. An overview of the algorithm is given as follows:

Algorithm 1 Global optimization method

1: Choose a number of initial evaluation points.
2: The objective function f is evaluated for all new members of the set.
3: A Kriging surrogate model is fitted to the values of the objective function.
4: Maximization of the expected improvement objective function $E[I]$.
5: The result of the maximization (the next input point most likely to improve the objective function) is added to the set.
6: The process repeats from step 2 until a predetermined number of iterations is reached or
$$\frac{MaxE[I(x)]}{f_{min}} < \epsilon.$$

5. Numerical Experiments and Results

5.1. Synthetic Numerical Results

In order to validate and to explain the Global Optimization method, we first made a test with a simple synthetic example . We consider the true function:

$$f(x) = x.\sin(x) + x.\cos(2x). \tag{7}$$

Our main goal is to find the global optimum of the true function (7) (the solid blue line in Figure 2). We consider that we have only few points generated by this true function (the red star in Figure 2), and we create an approximated response surface using the Kriging method (the black dotted line in Figure 2). It is associated with the standard error MSE (shown as the green line in Figure 2). Now, we apply the global optimization algorithm described above

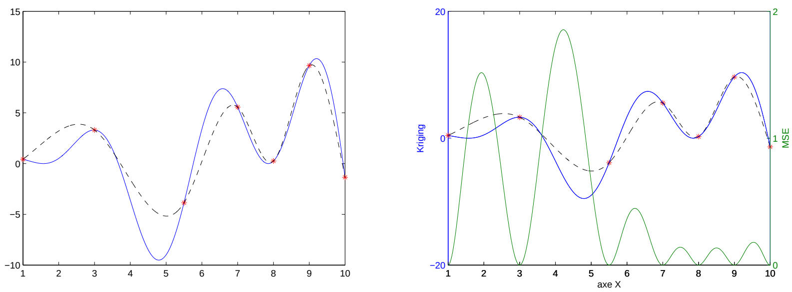

Figure 2: The true function (blue solid line), the set points used for the kriging interpolation (red star), the surface Kriging (black dashed line) and the estimator error (green solid line)

to the problem constructed by the set the real points $\Big(f(x_1), ..., f(x_7)\Big)$(the red star) and the surface $\tilde{f}(x)$ obtained by Kriging. In Figure 3, we present the expected improvement criterion $E[I(x)]$ (the green solid line in Figure 3) and, where it reaches its maximum value, we added another point in the set (now we have 8 real points). The $f(x)$ coordinate of this point is evaluated using the true function 7 (the black triangle in the right Figure 3). This process was repeated until the sampling point does not change and the global minimum of the objective function has been found. The global optimization method has run using an initial sampling of 7 points to build the surrogate (the shaded line in Figure 4). A further 5 function evaluations (the triangle in Figure 4) were

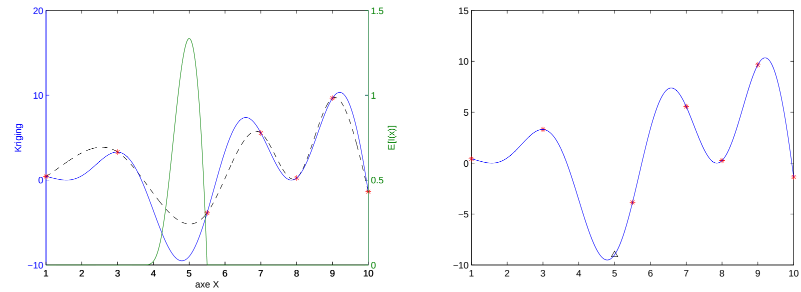

Figure 3: The true function (blue solid line), the set of points used for the kriging interpolation (red star), the Kriging surface (black dashed line), the Expected Improvement criterion (green sold line) and in the right the true function with the add point (the triangle)

required to find the global minimum. Now, we apply the global optimization method with an initial sampling of 2 points and the method is able to find reasonable solution in 13 function evaluations (see right of Figure 4).

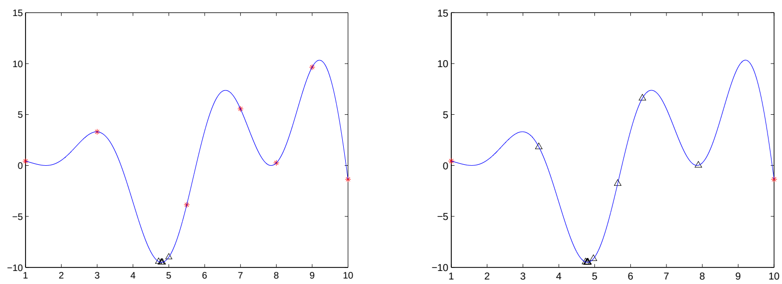

Figure 4: In the left the convergence method with 7 points and in the right the convergence with 2 points as initial set

5.2. Applications of Inverse Ellipsometric Problem

Now, we consider the inverse ellipsometric problem for thin film characterization. We are interested to find the shape defined in Figure 5, described by (H, CD) from measurements data. For this, we applied the global optimization algorithm using the Least Square cost function (4) from simulated data where (H, CD) parameters of optimization.

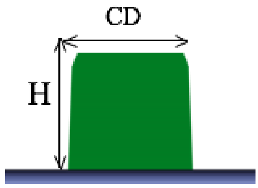

Figure 5: the pattern shape determining

First, we present the numerical solution of the direct problem $\Big(Is(t_i), Ic(t_i)\Big)$, $i = 1, ..., n$ with $H = 100nm$ and $CD = 150nm$. The latter is solved numerically with the Modal Method by Fourier Expansion[10] (Figure 6). To identify

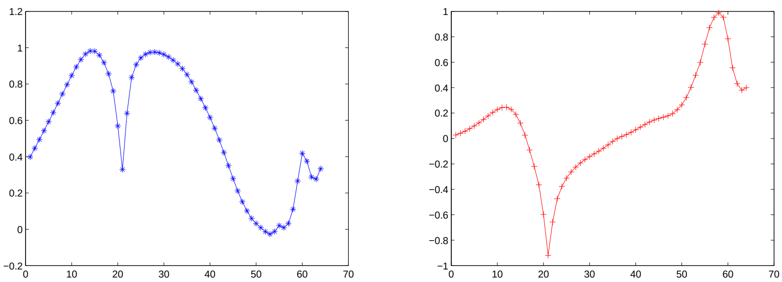

Figure 6: Solution of direct problem, in the right Is, in the left Ic with $H = 100nm$ and $CD = 150nm$

the parameters model (H, CD), we generate simulated test data using the solution of direct problem and we apply the algorithm described in Section 4.3. For the tests with noisy data, we perturbed the numerical solutions, by a Gaussian noise with fixed amplitude. In Figure 7, we present the convergence of global optimization where the initial evaluation points is 4 corners points (red star in the figure) and the new evaluation points (the triangle in the figure). The algorithm converges to $H = 99.5733$ and $CD = 150.2976$ after 31 evaluations of cost functions. In the following, we present the identification results by using the algorithm described in the paragraph 4.3 for different shapes. In the following table, the first column contains the exact shape (H, CD) from which we

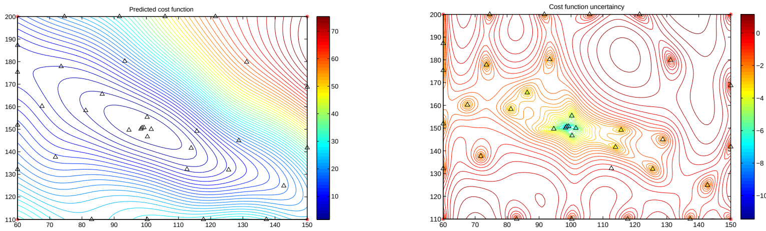

Figure 7: In the left, the evaluation of the objective function at the evaluation points, in the right, the maximum of the expected improvement $E(I)$

have formed measurements Ic^m, Is^m, the second column is reserved for identified shapes by optimization algorithm. Numerical tests have been done using the Gaussian noise fixed at 3%.

The exact shapes	Identified shapes
100 150	99.8432 150.1517
100 120	99.9457 119.8156
80 150	79.3635 150.6197
90 160	90.3090 159.6743

Table 1: Identification results of differents shapes

In what follow, we present some numerical tests which show the comparison between the noisy data and the solution of the direct problem generated by the identified parameters.

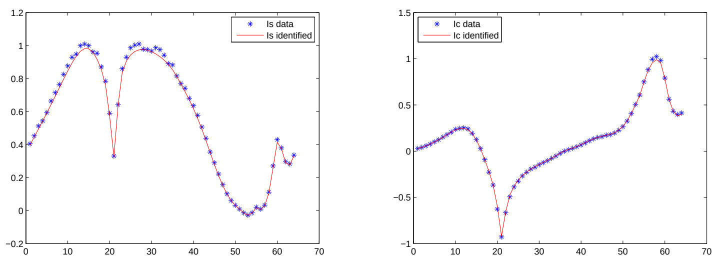

Figure 8: comparison between the noisy data and the solution of the direct problem generated by the identified parameters

5.3. Experimental Measurements

Now, we use experimental measurements delivered over the laboratory LTM. Then we applied the global optimization algorithm combined with the Kriging model.

5.3.1. Real Case: A Film Plan of Thickness

In order to find the thickness of a thin polymer film which optical indices n and k are known. We use the objective function described in section 3 by equation (4). The direct problem computes theoretically the ellipsometric signal values Is^t and Ic^t and the ellipsometer provides us with the measured data Is^m and Ic^m.

To show the performance of our method, we compared it with the conventional method of optimization (classical regression). Using a classical optimization scheme based on conjugate gradient method, we found the following results:

- If the initial point is 150, the method converges to $99,4861$ with 32 evaluations of the objective function.
- If the initial point is 170, the method converges to 318.6035 with 1474 evaluations of the objective function.

We remark that the final optimum provided by this method depends on the initial point and aimed at providing only a local minimum.

Now, we present the results of the global optimization method. We show in the next Figure 9, the initial previously known points and the point which were added during the EGO process. In Figure 9, we initiated the algorithm by 9 points (red star in Figure 9).

In this case, we obtained a convergence after 21 evaluations (blue plus in Figure 9). When using 6 points of initialization, the EGO algorithm converged after 26 evaluations which is much lower than the number of evaluations needed with the local optimization algorithm. In order to push the limits even further, we tried to use as few as 4 and even 2 values for initial points. The results of the EGO algorithm are reported in Figure 10. We can see that the EGO is able to find reasonable solution in 32 and 36 function evaluations. We present this results in the Table 2.

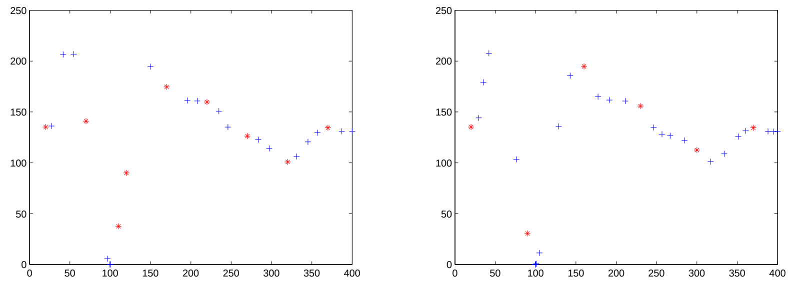

Figure 9: On the left, the EGO with 9 values for initial points, on the right, the EGO with 6 initial points

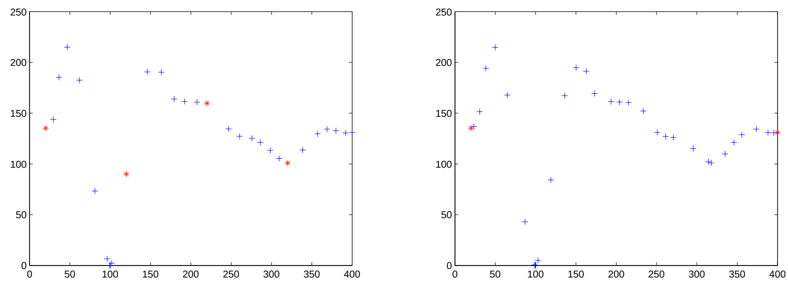

Figure 10: In the left, the EGO with 4 values for initial points, in the right, the EGO with 2 initial points.

Number measurements	Number evaluations
9	21
6	26
4	32
2	36

Table 2: Number of mesasurements and the objective function evaluation

Acknowledgments

The experiments measurements are delivered by the LTM-laboratory Grenoble, France, for which the author express his sincere gratitude. I also thank the anonymous referees for their valuable comments and help.

References

[1] L. Afraites, A. Bellouquid, Global optimization approaches to parameters identification in an immune competition model, *Commun. Appl. Ind. Math*, **5** (2014), 468-486.

[2] L. Afraites, J. Hazart, P. Schiavone, Application of the Kriging method to the reconstruction of ellipsometric signature, *Microelectronic Engineering*, **86**, No-s. 4-6 (2009), 1033-1035.

[3] R.M. Azzam, N.M. Bashara, *Ellipsometry and Polarized Light*, North-Holland (1977).

[4] V. Borboet al, On the solution of the incorrect inverse ellipsometric problem, In: *Proc. SPIE: 3485* (1998).

[5] N.A. Cressie, *Statistics for Spatial Data*, Wiley, New York (1993).

[6] Donald R. Jones, M. Schonlau, William J. Welch, Efficient global optimization of expensive black-box functions, *Journal of Global Optimization*, **13** (1998), 455-492.

[7] J. Martin, T. Simpson, Use of adaptive metamodeling for design optimization, In: *Proc. of the 9 AIAA/ISSMO Symposium on MAO, Atlanta* (2002), 2167-2175.

[8] J. Mockus, V. Tiesis, A. Zilinskas, The application of bayesian methods for seeking the extremum, In: *Towards Global Optimization*, North Holland, Amsterdam, **2** (1978), 117-129.

[9] M.A. Olive, R. Webster, Kriging: A method of interpolation for geographical information systems, *Int. J.of Geographical Information Systems*, **4**, No 3 (1990), 313-332.

[10] R. Quintanilha, Étude du problème inverse en diffractométrie spéctroscopique, *PhD Thesis*, INPG, Grenoble-France (2005).

[11] C.J. Raymond et al, *Proceeding of the SPIE*, **5375** (2004), 564-575.

[12] J. Sacks et al, Design and analysis of computer experiments, *Statistical Science*, **4**, No. 4 (1989), 409-423.

[13] Sébastien Soulan, *Développement de la Scatterométrie Dynamique pour le Suivi en Temps Réel de Procédés: Application la Microlectronique*, PhD Thesis, UJF, Grenoble-France (2008).

[14] Sébastien S. Besacier, T. Leveder, P. Schiavone, Real-time profile shape reconstruction using dynamic scatterometry, *Proc. SPIE*, **2618**, No. 1 (2007).