

# Characterization of optical diffraction gratings by use of a neural method

Stéphane Robert, Alain Mure-Ravaud, Dominique Lacour

► **To cite this version:**

Stéphane Robert, Alain Mure-Ravaud, Dominique Lacour. Characterization of optical diffraction gratings by use of a neural method. Journal of the Optical Society of America. A Optics, Image Science, and Vision, Optical Society of America, 2002, 19 (1), pp.24-32. 10.1364/JOSAA.19.000024 . ujm-02060281

**HAL Id: ujm-02060281**

**<https://hal-ujm.archives-ouvertes.fr/ujm-02060281>**

Submitted on 8 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Characterization of optical diffraction gratings by use of a neural method

Stéphane Robert and Alain Mure-Ravaud

*Laboratoire Traitement du Signal et Instrumentation, Unité Mixte de Recherche, Centre National de la Recherche Scientifique 5516, 23 rue du Docteur Paul Michelon, 42023 Saint-Etienne Cedex 2, France*

Dominique Lacour

*Laboratoire des Sciences et Matériaux pour l'Electronique et d'Automatique, Unité Mixte de Recherche, Centre National de la Recherche Scientifique 6602, 24 avenue des Landais, 63177 Aubière Cedex, France*

Optical scatterometry by use of a neural network is now recognized as an efficient method for retrieving dimensions of gratings in semiconductors or glasses. For an on-line control, a small number of measurements and a rapid data treatment are needed. We demonstrate that these requirements can be met by combining data preprocessing and a proper neural learning method. A good accuracy is attainable with the measurement of only a few orders, even in the presence of experimental errors, with a reduction in learning and computing time.

## 1. INTRODUCTION

The increasing integration density in integrated electronic circuits gives rise to new manufacturing capabilities that can be used in the optical domain for the realization of periodic structures of increased spatial frequency. Smaller grating periods lead to fewer diffracted orders and also to less-complex interferograms between overlapping orders. New application possibilities are currently offered by submicrometer diffraction gratings in rapidly growing fields such as optical communications and sensors.

A reliable manufacturing process for diffraction gratings can be controlled and possibly improved only if the critical dimensions can be accurately measured. The deterministic control of the process is ensured by the determination of the grating parameters that characterize the profile of the grooves. Two classical measurement methods are commonly used: scanning electron microscopy and atomic force microscopy. The first one is considered the main reference method in terms of accuracy. However, it requires some preparation for metal deposition in the case of dielectric gratings and etching. Furthermore, it is a destructive method usable only as an *a posteriori* sampling process control. The second one has an extreme sensitivity, but the dimension of the tips limits its spatial resolution and reproducibility. However, both methods perform a local measurement over a few periods only, require complex and costly equipment, and are very time-consuming. They are not adequate for the on-line control of the different steps of a manufacturing process.

Thus another type of characterization method is needed. The most adequate is based on measuring the scattered and/or the diffracted light of a grating under known illumination conditions and then on finding the profile of the structure from the measurement of dif-

fracted intensities. The last operation is known as the inverse-problem resolution. These two stages must be optimized for the low response time needed for a realistic industrial measurement method: The number of diffracted intensities to be measured must be sufficiently low and the treatment of the experimental data instantaneous. If these conditions are satisfied, it is a quick, non-contact, and nonlocal method. Furthermore, it requires low-cost optical equipment, and the measurement is easy to perform.

The theoretical problem of calculating the diffracted intensities in the case of wavelength-scale grating periods can be solved by a number of numerical techniques. For solving the inverse problem, theoretical and statistical methods have been tested. Theoretical methods developed so far<sup>1</sup> work only for some specific cases in particular conditions. Another method consists of retrieving the profile parameters by data analysis based on different statistical models, for instance, classical linear regression models, such as the principal components analysis (PCA) method,<sup>2</sup> the discriminant analysis method,<sup>3</sup> and the partial-least-square method.<sup>4-6</sup> In the past few years another statistical tool has received increased attention: the neural network. A neural network analysis can be regarded as a regression method based on a non-linear model. In the optical diffracted domain, a neural network has been used to develop an inspection method for a grating structure.<sup>7</sup> Thus several methods for solving the inverse problem include a neural network.<sup>8</sup> Some early results obtained by means of a neural analysis were not as accurate as those obtained from classical linear regression, such as the partial least square method<sup>4</sup> in particular. The accuracy of the results reported in more recent studies<sup>9,10</sup> is increasing, however, as a result of better comprehension of neural network function. Recently it

was shown in particular that neural networks resist noise more efficiently than does the partial least square method.<sup>11</sup> Nevertheless, all neural networks reported so far require a large number of inputs and neurons. This results in a highly complex network architecture and, above all, too long a training time.

In this paper we demonstrate that an equivalent accuracy is now attainable with neural networks of a lower level of complexity. It describes in detail every step of a neural network approach. The emphasis will be placed on a reduction of computer time by means of a reduction in complexity of the network and on a smaller number of input data so as to allow for a real-time measurement.

Section 2 describes the type of gratings considered and illustrates the complexity of the dependence of the diffraction characteristics on the profile parameters. The difficulty of choosing the best experimental condition is presented. The selected neural network structure is described in Section 3 with some general discussion about neural networks. The method for obtaining training samples is explained. Section 4 is concerned with the optimization of the learning step. The efficiency of PCA for reducing learning time is shown. In Section 5 the effect of experimental errors is considered. The increase of measurement errors indicates a poor generalization capacity. A solution is found in the use of noisy data in the learning procedure. An example of results obtained with different experimental conditions illustrates the efficiency of the method. The conclusion emphasizes the advantages and weaknesses of the method and identifies where future work should be directed.

## 2. INVERSE-SCATTERING PROBLEM

The type of grating profile considered in this work is a wavelength-scale period structure assumed to have a symmetric trapezoidal shape. This profile was chosen as a simulation of those usually obtained by photolithography transfer and subsequent reactive ion beam etching on a quartz substrate. The grating period is constant and assumed to be known. The characterization of this groove shape is thus made by the determination of three grating parameters: the sidewall projection  $b_1$ , the linewidth  $b_2$ , and the groove depth  $h$  (Fig. 1). All computer simulations are made with a fixed nominal grating period  $\Lambda = 1 \mu\text{m}$ . If not specified, the wavelength of the incident light is that of a laser diode at  $\lambda = 670 \text{ nm}$ . Hereafter the use of efficiencies will be preferred to intensities. They are defined as normalized intensities or equivalently as a percentage of the total incident intensity. The substrate being transparent, diffracted orders may be re-

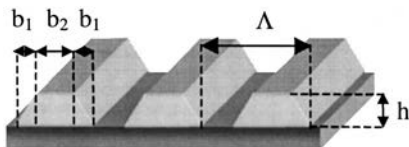


Fig. 1. Symmetric grating profile, defined by sidewall projection  $b_1$ , linewidth  $b_2$ , groove depth  $h$ , and period  $\Lambda$ .

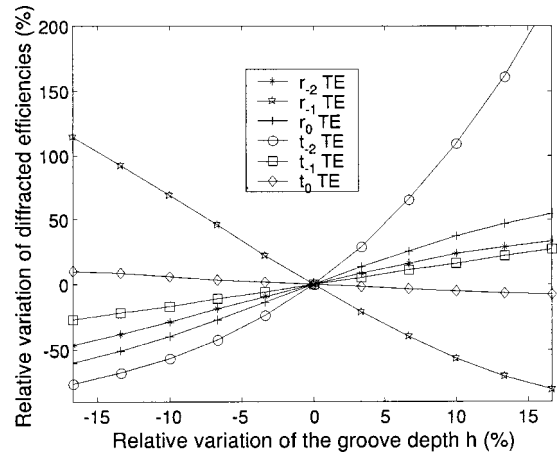


Fig. 2. Relative variations of diffracted efficiencies versus the groove depth. The grating parameters are  $b_1 = 0.105 \mu\text{m}$ ,  $b_2 = 0.32 \mu\text{m}$ , and  $h = 0.32 \mu\text{m}$ . Incident light is TE polarized at  $\theta = 30^\circ$ .

flected and transmitted. The reflected and transmitted efficiencies in the direction of the  $m$ th order will be denoted  $r_m$  and  $t_m$ .

Several rigorous numerical methods allow the calculation of the diffraction efficiencies in the different orders for a set of experimental conditions. They are defined by the angle  $\theta$ , the wavelength  $\lambda$ , and the TE or TM polarization of the incident light in a grating structure characterized by the refractive index of each medium and the groove parameters. A rigorous coupled-wave analysis was used here: the multi-layer modal method by Fourier expansion,<sup>12</sup> which gives accurate results whatever the polarization and grating period.

For gaining a more intuitive representation of the dependence of diffracted efficiencies on the grating-parameter variation, some preliminary calculations are performed. The diffraction efficiencies are computed versus the same groove-depth variation in a number of different structures under different incidence conditions. The efficiencies are calculated on the one hand for the same trapezoidal grating ( $b_1 = 0.105 \mu\text{m}$ ,  $b_2 = 0.32 \mu\text{m}$ , and  $h = 0.32 \mu\text{m}$ ) for two different angles of incidence ( $\theta = 30^\circ$  in Fig. 2 and  $\theta = 20^\circ$  in Fig. 3); on the other hand, for the same incidence angle  $\theta = 30^\circ$  for two different gratings ( $b_1 = 0.105 \mu\text{m}$ ,  $b_2 = 0.32 \mu\text{m}$  in Fig. 2 and  $b_1 = 0.00 \mu\text{m}$ ,  $b_2 = 0.50 \mu\text{m}$  in Fig. 4). The plot of the relative variation of efficiencies is preferred to reveal the sensitivity to the parameters. Figures 2–4 reveal that diffraction-order efficiencies are not equally sensitive to groove-depth variations. Some diffracted efficiencies are more sensitive (for example,  $t_{-2}$ ) than others (for example,  $r_0$ ) in Figs. 2 and 4. It can be noticed that the quantitative variations greatly change with parameters' values or incidence conditions; for example, in Fig. 3, the reflected order  $r_0$  is clearly more sensitive than the transmitted diffraction order  $t_{-2}$ . In addition, some attention must be paid to the absolute values of efficiencies that cannot be usable because of their weakness. For example, as shown in Fig. 5, the optimal diffracted-order efficiency  $t_{-2}$  cannot be easily measured with sufficient accuracy when the grating is illuminated under a  $30^\circ$  incidence angle.

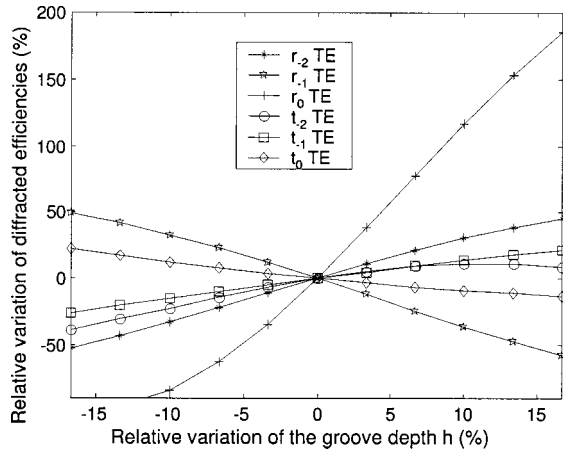


Fig. 3. Relative variations of diffracted efficiencies  $r_{-2}$ ,  $r_{-1}$ ,  $r_0$ ,  $t_{-2}$ ,  $t_{-1}$ , and  $t_0$  of a grating with  $\Lambda = 1 \mu\text{m}$ ,  $b_1 = 0.105 \mu\text{m}$ ,  $b_2 = 0.32 \mu\text{m}$ , and  $h = 0.32 \mu\text{m}$  when the groove depth is varying with TE polarized light at  $\theta = 20^\circ$ .

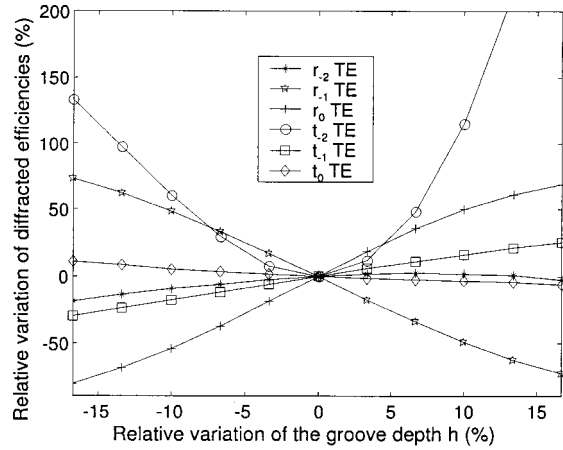


Fig. 4. Relative variations of diffracted efficiencies  $r_{-2}$ ,  $r_{-1}$ ,  $r_0$ ,  $t_{-2}$ ,  $t_{-1}$ , and  $t_0$  of a grating with  $\Lambda = 1 \mu\text{m}$ ,  $b_1 = 0 \mu\text{m}$ ,  $b_2 = 0.5 \mu\text{m}$ , and  $h = 0.32 \mu\text{m}$  when the groove depth is varying with TE polarized light at  $\theta = 30^\circ$ .

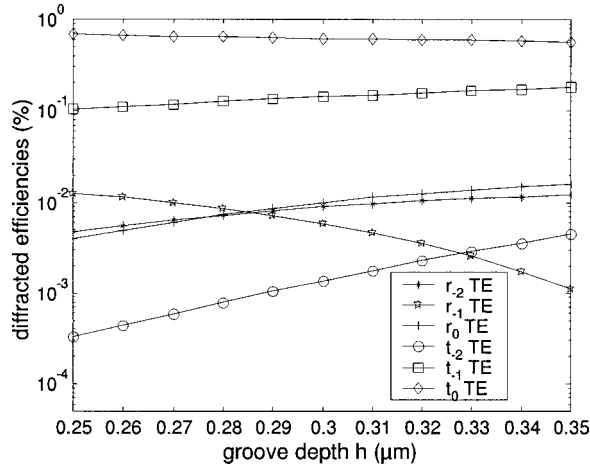


Fig. 5. Absolute variations of diffracted efficiencies  $r_{-2}$ ,  $r_{-1}$ ,  $r_0$ ,  $t_{-2}$ ,  $t_{-1}$ , and  $t_0$  of a grating  $\Lambda = 1 \mu\text{m}$ ,  $b_1 = 0.105 \mu\text{m}$ ,  $b_2 = 0.32 \mu\text{m}$ , and  $h = 0.32 \mu\text{m}$  when the groove depth is varying with TE polarized light at  $\theta = 20^\circ$ .

The diffracted efficiencies depend on grating parameters ( $b_1$ ,  $b_2$ , and  $h$ ) but also on incidence conditions such as angle  $\theta$ , polarization (TE or TM), and wavelength  $\lambda$ . Thus one must not make quick conclusions about the utility of a particular order on the basis of a single curve under specific conditions. However, in the case of a small grating period, the number of diffracted orders is small. Useful information will be obtained by the measurement of all high-intensity orders for various incidence conditions.

### 3. NEURAL NETWORKS

Neural networks can be used to solve problems that are difficult for conventional computing methods. They are based on biological neuron systems and have consequently an interesting capacity for learning.<sup>13</sup> In this work, a neural network will be trained to learn the relationship between diffraction efficiencies and grating parameters. The former will be the inputs and the latter the outputs of the network.

Neural networks used in this paper belong to the so-called multiple-layer structure. Each neuron in a layer is connected only to all those belonging to the previous layer or to the next layer. The present network contains three layers of neurons (Fig. 6): The first one is the input layer, the second one the hidden layer, and the third one the output layer. The organization of a simple neuron is represented in Fig. 7. The synaptic weight  $w_{i,j}$  is the strength of the connection between the output of neuron  $j$  (or input  $j$ ) and the input of neuron  $i$ . The sum  $a_i$  of weighted inputs is the argument of the transfer function  $f$ . The output  $o_i$  of the neuron  $i$  is thus obtained by  $f(a_i)$ .<sup>13</sup> A sigmoid transfer function was selected for neu-

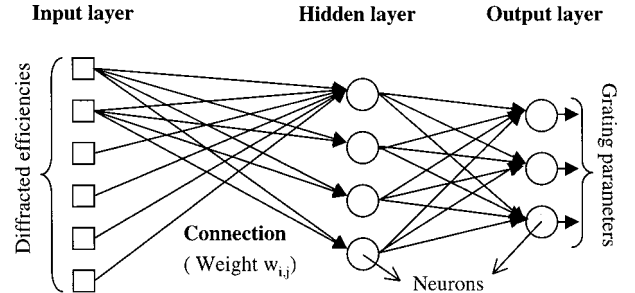


Fig. 6. Graph of a three-layer network with six inputs, three outputs, and four neurons in the hidden layer.  $w_{i,j}$  is the connection weight between the  $j$ th neuron in the input layer and the  $i$ th in the hidden layer.

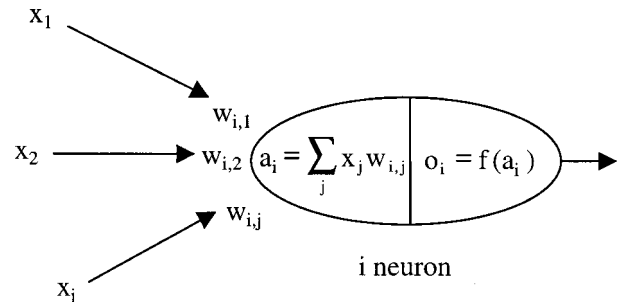


Fig. 7. Structure of a formal neuron.  $x_j$  represents one input of the neuron,  $a_i$  the weighting sum and  $o_i$  the neuron output.

rons in the hidden layer and a linear transfer function for the output layer. It was demonstrated that this kind of neural network, with these transfer functions, is able to approximate any function with a finite number of discontinuities.<sup>14,15</sup> It will then be suitable for solving the inverse problem.

The use of neural networks is divided into two steps: the learning step and the exploitation step. The learning step needs a representative set of data pairs that will be called {input/target} couples. It consists of an adjustment of the synaptic weights. The initialization of these values is ensured by the Nguyen–Widrow algorithm.<sup>16</sup> This adjustment is performed by an iterative procedure to minimize the error between the target and the calculated outputs of the network. The {input/target} couples from the training data set are randomly provided to the network, and outputs calculated with previous weights  $w$  are compared with theoretical targets for each sample. Thus the estimated error allows the adjustment of the connection values between neurons by a gradient descent. The faster backpropagation Levenberg–Marquardt training algorithm<sup>17</sup> was used for this training process. Weights are updated after each epoch, i.e., the presentation of all training couples. This phase is named “batch supervised training section.” This operation is repeated until a definite desired value of the calculated output error has been reached and can be performed within a few minutes.

The second operation can be called the exploitation step. Synaptic weights are now fixed. New unknown inputs are introduced into the network. Outputs are instantaneously calculated, giving an estimation of the grating parameters.

In a real characterization experiment the set of training samples will be provided by a representative selection of experimental data. The main problem occurring in the implementation of this method is that neural network training often requires nearly 1000 samples. The generation of such a number of data is intractable, and above all, the accurate reproducible nonlocal measurement of the parameters is not realistic. Instead, the numerical simulation of various diffraction gratings is easy, and diffracted efficiencies can be accurately obtained for known parameters. The simulation will give us a set of samples called exact values. Furthermore, this method has no limit in resolution. One assumption to be verified is the symmetric trapezoidal waveform of the grating profile. It must be remembered that neural network results are valid only within the training domain. This assumes that the grating-parameter variations are approximately known.

## 4. IMPLEMENTATION OF THE METHOD

### A. Theoretical Samples

Grating fabrication leads to deviations of the different grating parameters relative to their targeted values. The simulated structure is a transparent grating made by dry etching in a quartz substrate (1.456 index at the used wavelength). Its targeted geometrical parameters are  $b_1 = 0 \mu\text{m}$ ,  $b_2 = 0.5 \mu\text{m}$ , and  $h = 0.3 \mu\text{m}$ . It is assumed that the actual parameter values obtained after the fabrication process are kept within the ranges 0

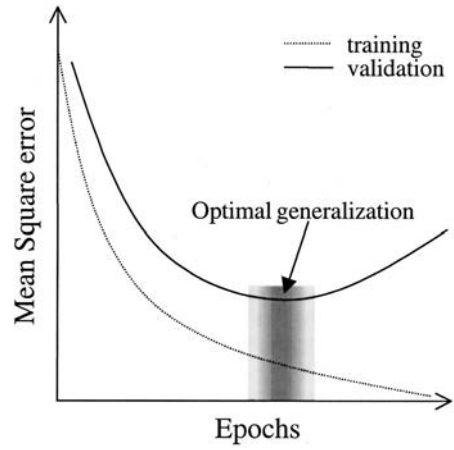


Fig. 8. Typical behavior of the rms error on training and test data sets during the training process.

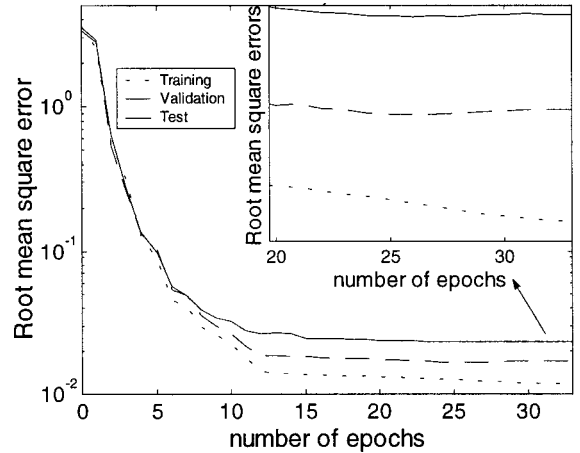


Fig. 9. Experimental variation of the error during the training process, for a neural network composed of 18 inputs, 3 outputs, and 15 neurons in the hidden layer. The training requires 1300 couples. It is stopped when the error calculated on the validation set reaches a minimum (33 epochs).

$< b_1 < 0.15 \mu\text{m}$ ,  $0.2 \mu\text{m} < b_2 < 0.5 \mu\text{m}$ , and  $0.25 \mu\text{m} < h < 0.35 \mu\text{m}$ . There are 18 inputs to the neural network ( $r_0$ ,  $t_0$ , and  $t_{-1}$  efficiencies in TE and TM polarized light for three values of the incident angle:  $20^\circ$ ,  $25^\circ$ , and  $30^\circ$ ). A set of 1300 theoretical {efficiencies/grating parameters} couples that belong to the parameter-variation domain are calculated. The calculation time is quite long (several hours), but it has to be performed only once.

### B. Data Preprocessing and Postprocessing

Before neural treatment, all inputs and targets are normalized so that they have zero mean and unity standard deviation. Thus all variables will have the same influence during the neural training, whatever their size. A large number of inputs requires a larger neural network with a consequent increase in training-time requirements and the need for a greater number of training samples. The number of input vectors is still large (18), but they are probably correlated. A PCA was performed before the efficiencies to the network were provided. The PCA orthogonalizes the components of the input vectors and orders the resulting orthogonal components (principal

components) so that the largest variation is concentrated in a few components. The principal components that contribute to less than 1% in the variation of the data set were eliminated (i.e., all components for which the eigenvalue represents less than 1% of the total inertia). These

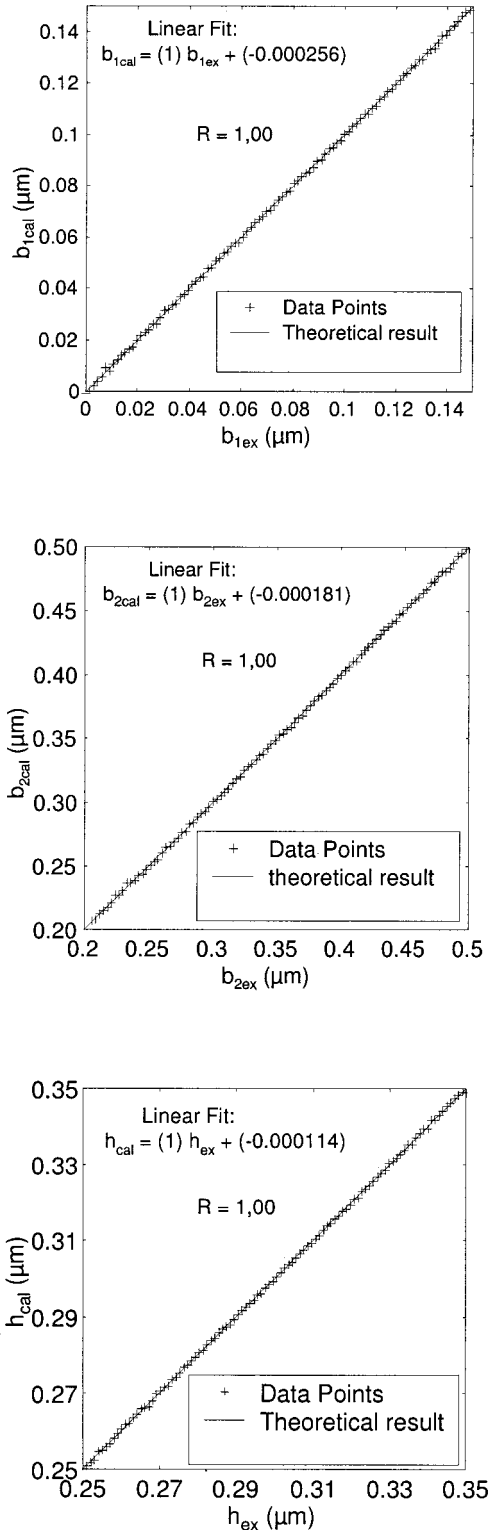


Fig. 10. Representation of each grating parameter calculated by a neural network with exact inputs versus its theoretical values. The training is performed with 1300 exact {input/target} calculated couples.

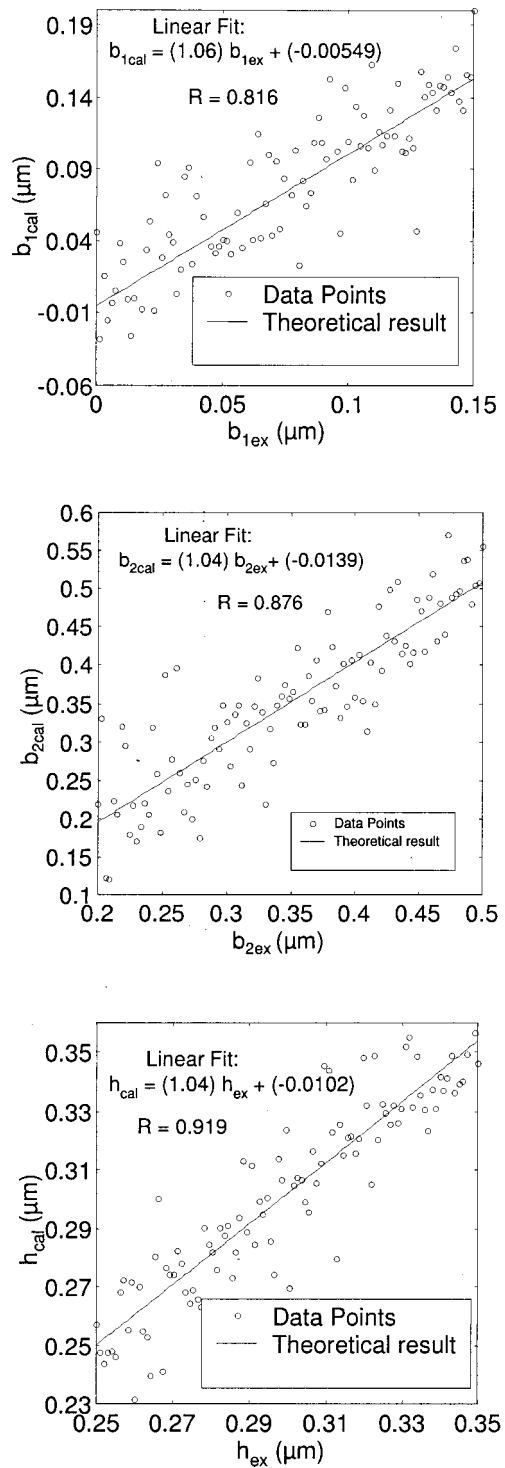


Fig. 11. Representation of each grating parameter calculated by a neural network with experimental simulation inputs versus its theoretical values. The training is performed with 1300 exact {input/target} calculated couples.

two data preprocessing steps make the convergence of the algorithm easier and therefore reduce the duration of the training. In our case, the input number of the network is reduced to 14. Thus training duration decreases from 4600 to 420 s.

A postprocessing step on the output data allows a scaling back to their values in the initial range.

### C. Particular Training: Early Stopping

One problem in using a neural network is to stop the training adequately. Indeed, owing to the training algorithm, output errors on training data may reach a value as small as desired through an increase in the number of epochs. Greater errors can then occur for unknown samples. The network loses its capacity to generalize new situations. Thus the {input/target} data set is ran-

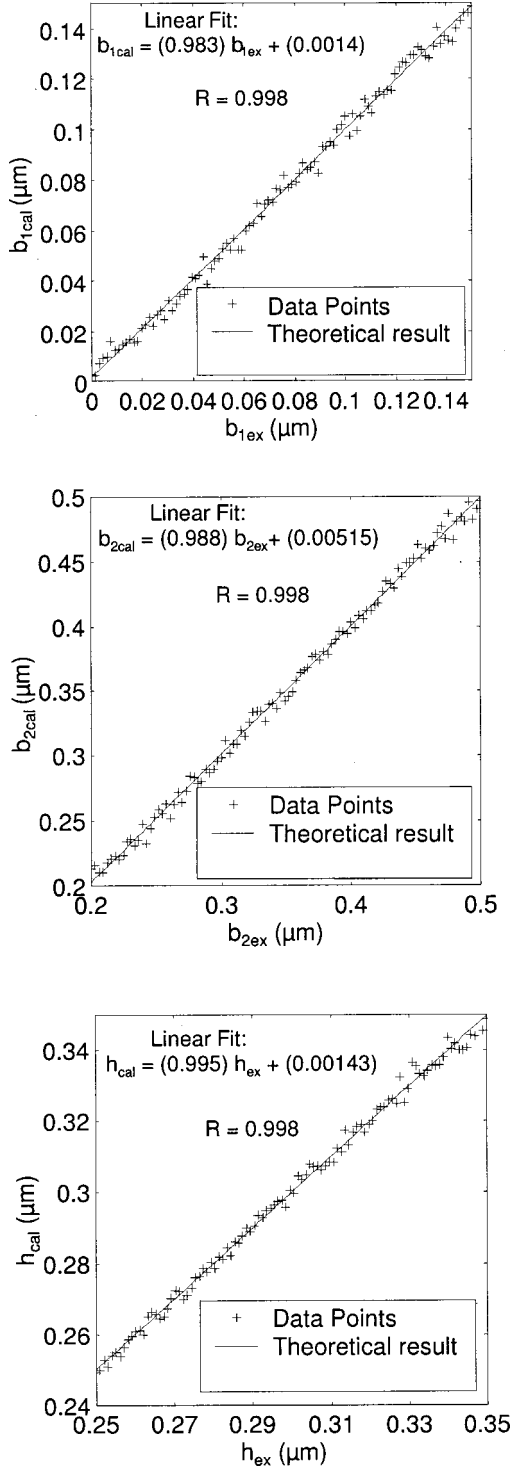


Fig. 12. Representation of each grating parameter calculated by a neural network with exact inputs versus its theoretical value. The training is performed with 1300 noisy {input/target} couples.

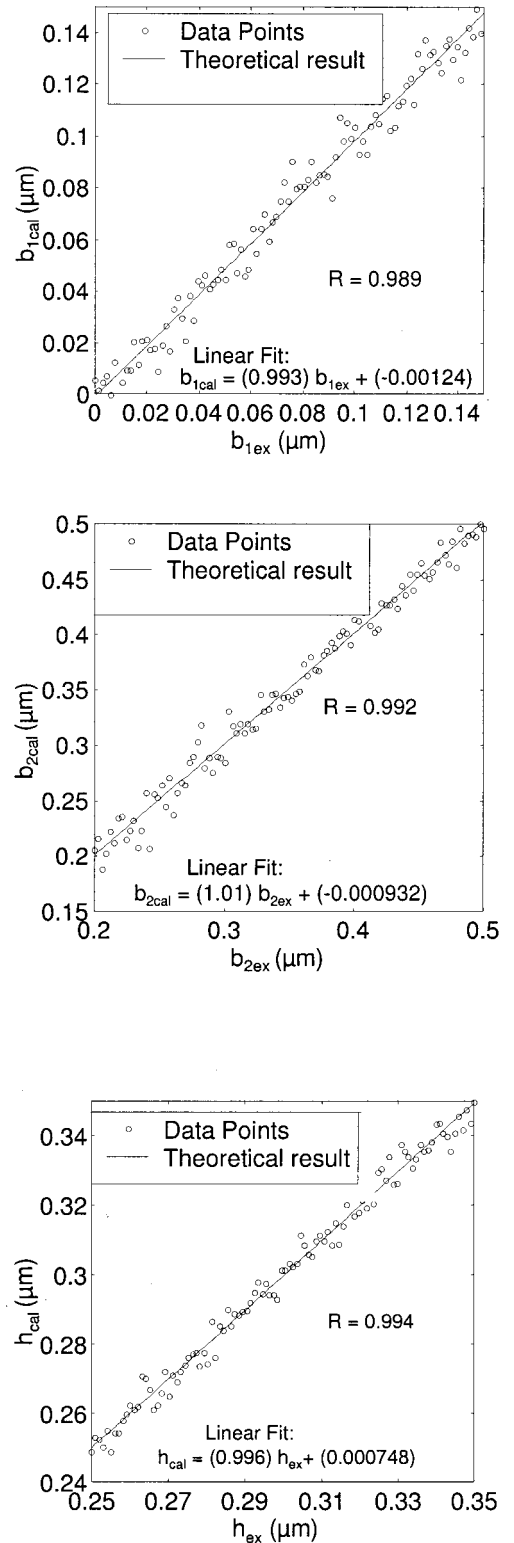


Fig. 13. Representation of each grating parameter calculated by a neural network with experimental simulation inputs versus its theoretical value. The training is performed with 1300 noisy {input/target} couples.

domly split into three parts to avoid this overfitting: the training, test, and validation data set. The first set, which is the largest, is used for weight adjustment; the second set, in stopping the training process; and the last set, in controlling the correct operation of the process.

Thus at each epoch, after updating the weights by forwarding the data from the training set and applying the Levenberg–Marquardt algorithm, samples from the test set are provided to the network and a test error is calculated. While the number of epochs is growing, a minimum in this error evolution can be observed, whereas the calculated error on the training data set is still decreasing. This typical variation is depicted schematically in Fig. 8. Thus the training must be stopped at this point to produce a good generalization. The errors calculated on the validation set must have the same behavior as the test error; otherwise, the indication is that an unsuitable division of the available data was made. This method is called early stopping. Figure 9 represents the error variations calculated on the different data sets when the training is performed by a neural network involving 15 neurons in the hidden layer. The training is stopped at the 26th epoch, corresponding to the maximum generalization capability.

## 5. RESULTS AND DISCUSSION

The neural network performances were evaluated on the basis of 100 new simulated gratings with parameters randomly chosen in the range defined above. The rms error calculated on each grating parameter  $p$  is defined by

$$e(p)_{rms} = \left[ (1/n) \sum (p_{ex} - p_{cal})^2 \right]^{1/2}, \quad (1)$$

where  $n$  is the number of simulated gratings ( $n = 100$ ),  $p_{ex}$  is the exact grating parameter value, and  $p_{cal}$  is the calculated value given by the neural network.

At first, a simulation is performed with these samples by the neural network trained as described earlier. The plots of neural calculated values  $p_{cal}$  versus exact ones  $p_{ex}$  are represented in Fig. 10. A linear regression analysis is performed between them to quantify the estimation validity. The linear fit is practically identical to the perfect one; i.e, its slope is near 1, the correlated coefficient  $R$

**Table 1. Summary of the rms Errors Calculated for Different Training and Test Sets**

Training Sets	Test Sets	rms Error (nm)		
		$b_1$	$b_2$	$h$
Training with exact values	Theoretical inputs	0.56	1.03	0.32
	Noisy inputs	27.4	43.81	12.96
Training with noisy values	Theoretical inputs	2.7	4.97	1.3
	Noisy inputs	7.39	11.81	3.37

**Table 2. Effects of the Preprocessing (PCA) in the Reduction of the Input Number and Training Duration Time for Two Cases**

Training Sets	Without PCA		With PCA	
	Number of Inputs	Training Time	Number of Inputs	Training Time
Training with exact values	18	4635	8	320
Training with noisy values	18	450	14	300

= 1.00, and the  $y$  intercept near 0. The different rms errors for  $b_1$ ,  $b_2$ , and  $h$  are 0.56 nm, 1.03 nm, and 0.32 nm, respectively.

Nevertheless, it is important to test the method in an experimental condition. Adding a Gaussian noise to the exact diffracted efficiencies as follows makes a simulation of measured values,

$$\eta_n = \eta_{ex} - N(0,0.03), \quad (2)$$

where  $\eta_n$  is the noisy diffracted efficiency,  $\eta_{ex}$  the exact one, and  $N(0,0.03)$  is the normal distribution with a zero mean and a standard deviation equal to 3% of the diffraction efficiency. The results shown in Fig. 11 are clearly less accurate than those obtained in Fig. 10. The rms errors on grating parameters  $b_1$ ,  $b_2$ , and  $h$  rise to 27.4, 43.81, and 12.96 nm, respectively. The best linear fit is now different from the perfect one. The growth of the rms error is due to the introduction of simulated experimental errors with a Gaussian noise. This results from the fact that theoretical couples ensure the training, whereas the test is performed on noisy inputs. Thus the neural network is not trained on the data set corresponding to the experimental simulation. The solution to this problem will be to train the network with real gratings. But as stated above, the fabrication of a large number of gratings requires a very long time, and, above all, it is impossible to determine the grating parameters with a sufficient accuracy.

Another training of the neural network was then performed by using a noisy data set deduced from exact data. The same investigations were then made with exact inputs on the one hand (Fig. 12) and with experimental simulated inputs on the other hand (Fig. 13). Linear fits are nearly the same for both figures. All the training rms errors and computation time are summarized in Table 1 and Table 2, respectively. The errors on the theoretical input increase, whereas the error on experimental simulations decreases to reach nearly the same value (5–10 nm). Moreover, an interesting effect is observed: The training time decreases and reaches an acceptable value (a few minutes). These results show that the neural network can preserve accurate predictions in experimental simulation.

Finally, another type of validation of the proposed method is presented: One grating was randomly chosen:  $b_1 = 0.094 \mu\text{m}$ ,  $b_2 = 0.473 \mu\text{m}$ , and  $h = 0.260 \mu\text{m}$ . Table 3 gives for this sample the grating parameters calculated by the noisy training network from exact and noisy inputs. Next, the transmitted zero-order  $t_0$  efficiency is computed by the multilayer modal method by Fourier expansion for these two different reconstructed gratings. Figure 14 and Fig. 15 represent the efficiency



**Table 3. Grating Parameter Prediction for Two Kinds of Neural Inputs for a Randomly Chosen Grating**

Neural Inputs	Calculated Grating Parameters (nm)		
	$b_1$	$b_2$	$h$
Theoretical	93.94	472.73	260.1
Calculated with theoretical inputs	95.77	465.39	260.35
Calculated with noisy inputs	95.82	464.36	261.48

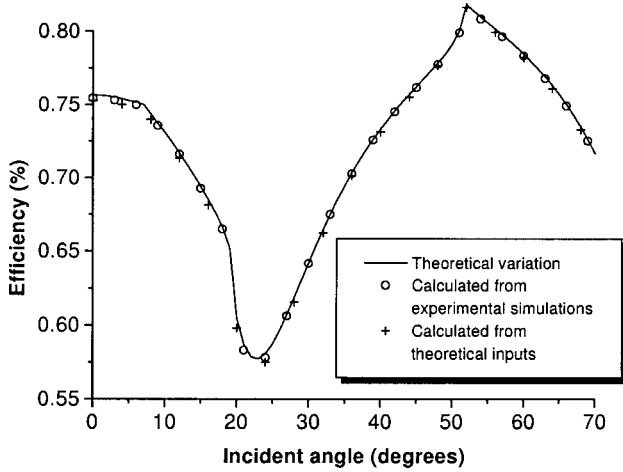


Fig. 14. Variation of the zero-order transmitted efficiency versus the incident angle at 670 nm wavelength calculated in a grating defined by  $b_1 = 0.094 \mu\text{m}$ ,  $b_2 = 0.473 \mu\text{m}$ , and  $h = 0.260 \mu\text{m}$ . The same efficiency variation calculated in the gratings reconstructed by the neural network for the two cases of inputs (theoretical and noisy inputs) is represented.

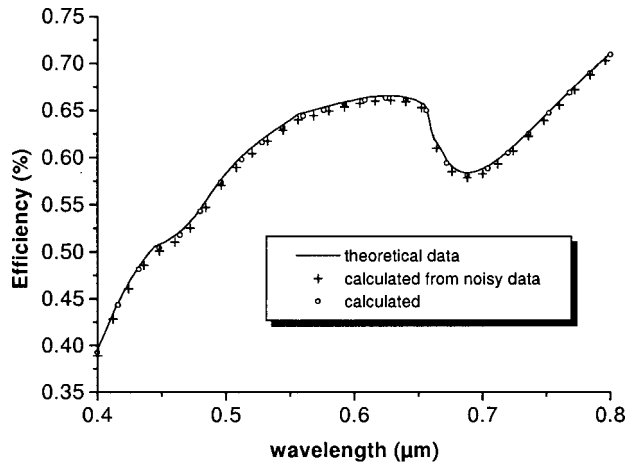


Fig. 15. Variation of zero-order transmitted efficiency versus wavelength at  $\theta = 20^\circ$  calculated in a grating defined by  $b_1 = 0.094 \mu\text{m}$ ,  $b_2 = 0.473 \mu\text{m}$  and  $h = 0.260 \mu\text{m}$ . The same efficiency variation calculated in the gratings reconstructed by the neural network for the two cases of inputs (theoretical and noisy inputs) is represented.

variation on varying the incident angle and the wavelength of incident light, respectively. Deviations from theoretical curves are small even if the incident angle and

the wavelength are far away from those used in the characterizing process. This allows us to draw conclusions on the validation of the method in any experimental conditions.

## 6. CONCLUSION

This paper developed the possibilities offered by multiple-layer neural networks to solve the inverse scattering problem applied to the characterization of gratings. The proposed method is closer to the operating conditions of a grating than are local characterization methods, for example, atomic force microscopy, because it achieves an average over thousands of groove profiles in the region where the beam impinges on the grating. The classical PCA used as a preprocessing of data has shown a very efficient effect in terms of the learning time of the network. This results in a significant decrease in the number of network inputs but unfortunately not in intensity measurements. Neural training with noisy inputs permits us to perform experimental estimation of parameters with good accuracy. Furthermore, the learning method based on theoretically calculated samples implies no limit in resolution. Nevertheless, it must be pointed out that it is not a real profile measurement but a grating modelization in the sense of the best fit with a trapezoidal shape giving the nearest diffracted intensities.

In our future research we will search for the most relevant diffracted orders and their experimental conditions in order to achieve a selection among all possible measurements. The aim is to reduce the necessary number of measured intensities which will be an important saving of time. Comparison with real experimental results is also one of our objectives.

## ACKNOWLEDGMENT

The authors are grateful to O. Parriaux for stimulating discussions leading to a more comprehensive presentation of this work.

Stephane Robert's e-mail address is stephane.robert@univ-st-etienne.fr.

## REFERENCES

1. A. Roger and D. Maystre, "Inverse scattering method in electromagnetic optics: application to diffraction gratings," *J. Opt. Soc. Am.* **70**, 1483–1495 (1980).
2. K. P. Giapas, R. A. Gottscho, L. A. Clark, J. Kruskal, D. Lambert, A. Kornblit, and D. Sinatore, "Use of light scattering in characterizing reactively ion etched profiles," *J. Vac. Sci. Technol. A* **9**, 664–668 (1991).
3. R. Krukar, A. Kornblit, L. A. Clark, J. Kruskal, D. Lambert, E. A. Reitman, and R. A. Gottscho, "Reactive ion etching profile and depth characterization using statistical and neural analysis of light scattering data," *J. Appl. Phys.* **74**, 3698–3706 (1993).
4. S. S. H. Naqvi, R. H. Krukar, J. R. McNeil, J. E. Franke, T. M. Niemczyk, D. M. Haaland, R. A. Gottscho, and A. Kornblit, "Etch-depth estimation of large-period silicon gratings with multivariate calibration of rigorously simulated diffraction profiles," *J. Opt. Soc. Am. A* **11**, 2485–2493 (1994).
5. R. H. Krukar, S. L. Prins, D. M. Krukar, G. A. Peterson, S. M. Gaspar, J. R. McNeil, and S. S. H. Naqvi, "Using scattered light modeling for semiconductor critical dimension metrology and calibration," in *Integrated Circuit Metrology*,

- Inspection, and Process Control VII, M. T. Postek, ed., Proc. SPIE **1926**, 60–71 (1993).
6. J. Bischoff, J. W. Baumgart, H. Truckenbrodt, and J. J. Bauer, “Photoresist metrology based on light scattering,” in *Metrology, Inspection, and Process Control for Microlithography X*, S. K. Jones, ed., Proc. SPIE **2725**, 678–689 (1996).
  7. A. D. Mc Aulay and J. Wang, “Optical diffraction of periodic structures using neural networks,” *Opt. Eng.* **37**, 884–888 (1998).
  8. J. N. Hwang, C. H. Chan, and R. J. Marks, “Frequency selective surface design based on iterative inversion of neural networks,” Presented at International Joint Conference on Neural Networks, Washington, D.C., 1990.
  9. I. Kallioniemi, J. Saarinen, and E. Oja, “Optical scatterometry of subwavelength diffraction gratings: neural network approach,” *Appl. Opt.* **37**, 5830–5835 (1998).
  10. I. Kallioniemi, J. Saarinen, and E. Oja, “Characterization of diffraction gratings in a rigorous domain with optical scatterometry: hierarchical neural-network model,” *Appl. Opt.* **38**, 5920–5930 (1999).
  11. J. Bischoff, J. Bauer, U. Haak, L. Hutschenreuther, and H. Truckenbrodt, “Optical Scatterometry of quarter-micron patterns using neural regression,” in *Metrology, Inspection, and Process Control for Microlithography XII*, B. Singh, ed., Proc. SPIE **3332**, 526–537 (1998).
  12. L. Li, “Multilayer modal method for diffraction gratings of arbitrary profile, depth, and permittivity,” *J. Opt. Soc. Am. A* **10**, 2581–2591 (1993).
  13. J. Hérault and C. Jutten, *Réseaux Neuronaux et Traitement du Signal* (Editions Hermes, Paris, 1994).
  14. G. Cybenko, “Approximation by superpositions of sigmoidal functions,” *Math. Control Signal Syst.* **2**, 303–314 (1989).
  15. K. I. Funahashi, “On the approximate realization of continuous mappings by neural networks,” *Neural Networks* **2**, 183–192 (1989).
  16. D. Nguyen and B. Widrow, “The truck back-upper: an example of self-learning in neural networks,” in *Neural Networks for Robotics and Control*, W. T. Miller, R. Sutton, and P. Werbos, eds. (MIT Press, Cambridge, Mass., 1990), Vol. 12, pp. 287–299.
  17. M. T. Hagan and M. Menhaj, “Training feedforward networks with the Marquardt algorithm,” *IEEE Trans. Neural Netw.* **5**, 989–993 (1994).