ELSEVIER

Information Sciences 129 (2000) 31–44

INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL

www.elsevier.com/locate/ins

Neural networks for HREM image analysis

Holger Kirschner, Reinald Hillebrand *

Max Planck Institute of Microstructure Physics, Weinberg 2, D-06120, Halle/Saale, Germany

Received 1 January 2000; received in revised form 4 July 2000; accepted 10 September 2000

Abstract

We present a new neural network-based method of image processing for determining the local composition and thickness of III–V semiconductors in high resolution electron microscope images. This is of great practical interest as these parameters influence the electrical properties of the semiconductor. Neural networks suppress correlated noise from amorphous object covering and distinguish between variations of sample thickness and semiconductor composition. © 2000 Elsevier Science Inc. All rights reserved.

Keywords: Neural network; Image processing; Electron microscopy; Compound semiconductor

1. Introduction

Imaging techniques and image processing methods play a central role in natural sciences. In particular, high resolution transmission electron microscopy (HREM) provides submicron information in physics and materials science. To quantify essential features of semiconducting materials, a neural network-based image processing approach has been elaborated. III–V semiconductor devices with systematically varied composition, so-called heterostructures, are of great practical interest. Nowadays, devices with such heterostructures are for instance, laser diodes and other quantum well structures. Typical material systems are:

* Corresponding author. Tel.: +49-345-5582911; fax: +49-345-5511223.
E-mail address: hi@mpi-halle.de (R. Hillebrand).

0020-0255/00/$ - see front matter © 2000 Elsevier Science Inc. All rights reserved.
PII: S 0 0 2 0 - 0 2 5 5 (0 0) 0 0 0 6 7 - 0

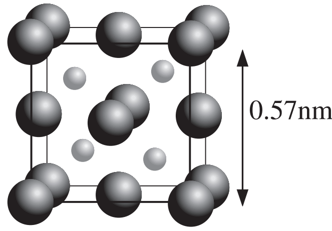

Fig. 1. Sphalerite structure unit cell: the two different sizes of spheres mark the two sublattices, e.g., Ga and As.

$$In_{1-x}Ga_xAs \quad \text{and} \quad Al_{1-x}Ga_xAs \tag{1}$$

where composition x varies in the range of [0, 1]. Such crystals are of sphalerite structure with a lattice parameter of about 0.5 nm (see Fig. 1). The sphalerite structure consists of two shifted fcc sublattices [1]. For physical reasons, the composition of one sublattice is varied in the crystal growth process (i.e., two elements statistically occupy the sites of one sublattice, while the other sublattice is homogeneous), which is also the case in examples (1). The best spatial resolution of composition determination methods is achieved by applying image processing to HREM images [2–4]. We present a method for determining composition and thickness from HREM images using neural networks. It should be noted here that alternative fuzzy logic approaches have also been elaborated and successfully applied to composition determination [5–9].

The method described here achieves a spatial resolution of about unit cell size (e.g., AlGaAs: 0.57 nm). Composition determination has to map a part of the image (image cell of N pixels, equals to a sample region of unit cell size) to a one-dimensional composition parameter x (cf. (1)). This is done in two steps:

$$\mathbb{R}^N \xrightarrow{p} \mathbb{R}^3 \xrightarrow{f} x \qquad x \in [0, 1]. \tag{2}$$

- image preprocessing p, which maps each image cell to a three-dimensional real vector using prior knowledge of crystal symmetry and imaging process (Section 2);
- approximation of function [1] f using neural networks (Section 4).

2. Image preprocessing

We cut the HREM image into sections which correspond to sample regions of unit cell size. The left column of Fig. 2 shows two examples (AlAs,GaAs) for

[1] Function f is only defined on a small subset of $\mathbb{R}^3$.

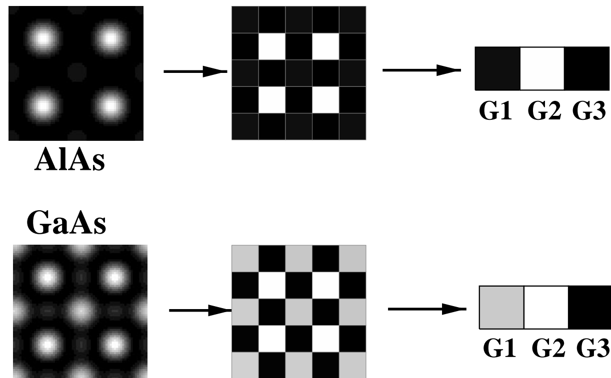

Fig. 2. Image preprocessing for simulated examples: GaAs and AlAs. First column: simulated images of unit cell size, second column: 25 values after evaluating the brightness, third column: three-dimensional *get3* vector.

1	3	1	3	1
3	2	3	2	3
1	3	1	3	1
3	2	3	2	3
1	3	1	3	1

Fig. 3. Second step of image preprocessing: the numbers show the equivalent positions for the *get3* averaging.

such regions. For each such part there are due to crystal symmetry only 25 sites where maxima or minima of brightness can appear (for detailed discussion see [10]). The brightness in these sites is evaluated by fitting rotational paraboloids of fourth-order to the image (The fourth-order approximation turned out to be superior compared to second-order and higher-order approximation). The second column of Fig. 2 shows the result of that first step for two simulated images.

According to the crystal point symmetry, we can identify three groups of equivalent positions, as shown in Fig. 3. Averaging over each group leads to a three-dimensional vector. In the following, we will call this vector *get3* (see Fig. 2, right).

3. HREM images

To get the function f in (2) between the *get3* vector and the composition x of the sample, we have to look closer to the nature of HREM images.

3.1. Template regions

Typical HREM images include regions where the composition and sample thickness are nearly constant by crystal growth. In the following, we will call these regions "template regions". Fig. 4 shows the HREM image of an AlGaAs sample where such template regions are marked. After image preprocessing we average for each template region over the included *get3* vectors. The results are two average experimental *get3* vectors and with that two experimental points of function f in (2).

3.2. Simulated HREM images

To interpolate the two experimental points of our desired function f we need to simulate *get3* vectors for certain ranges of sample composition, thickness and imaging conditions. We get these simulated *get3* vectors by simulating HREM images and performing *get3* image preprocessing on the simulated images analogous to the evaluation of experimental images.

For HREM image simulation, we use the EMS software package from Stadelmann [11,12]. This software package calculates dynamical electron diffraction by the multislice method. Images are calculated with nonlinear imaging theory. EMS is nowadays the most extensively tested and most accepted among HREM image simulation software.

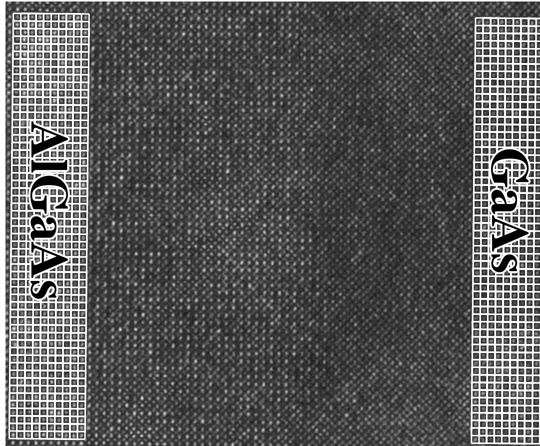

Fig. 4. HREM image of an AlGaAs sample with marked template regions at the right rsp. left boundary of the image.

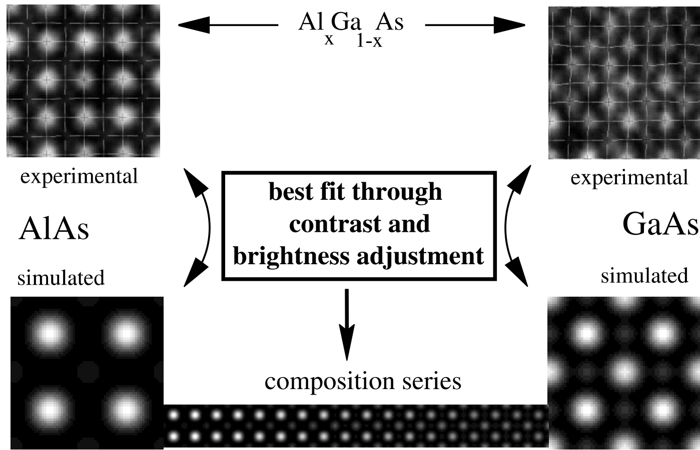

Fig. 5. Adjustment of brightness and contrast.

3.3. Comparing experiment and simulation

Before comparing the experimental *get3* vectors with simulated ones, it is necessary to adjust brightness and contrast of the simulation (i.e. average and standard deviation of the image intensities).

Fig. 5 illustrates the adjustment process for two template regions AlAs (left column) and GaAs (right column). Adjusting the brightness and contrast means to calculate the resulting image $\vec{R}$ from the raw image $\vec{I}$ as:

$$R_{ij} = b + aI_{ij} \quad a, b \in \mathbb{R}. \tag{3}$$

We want to find the two adjustment parameters a, b so that the simulated template images (left and right ends of bottom row in Fig. 5) are matching optimally the experimental ones (top row). This can be achieved by doing a least-squares fit [2] to get the best fitting of experiment and simulation.

3.4. Average experimental parameters

To get average experimental parameters for the template regions we compare the average experimental *get3* vectors with linear adjusted simulations varying the parameters of the simulation systematically. If we consider a

[2] The fitting process includes the constraint that only positive contrast adjustment is possible. Otherwise, we would consider image and inverse very similar, which has no physical reason, however.

HREM image with two template regions (as in 4) the simulation includes the following parameters:

$$\begin{array}{lll} \text{composition of the two template regions:} & x_1 & x_2, \\ \text{sample thickness of the two regions:} & t_1 & t_2, \\ \text{defocus of the electron microscope objective:} & \Delta. & \end{array} \quad (4)$$

The left-hand side of Fig. 6 shows the square deviation (dark $\hat{=}$ small deviation) of different simulations to one experimental image (two template regions). The ordinate corresponds to the thickness of one simulated template region t_1 and the abscissa corresponds to the defocus Δ in the simulated imaging process. It has to be noted that the defocus is an electron optical parameter which controls the contrast of the image. Δ is chosen >0 for contrast reasons [13]). For the other parameters in (4), the optimum values (minimum square deviation) are depicted.

To decide which of the combinations of experimental parameters have to be taken into account, we need to introduce an error limit. Simulations which exceed this error limit are not considered. A low boundary for choosing the error limit is the error in the experimental averages:

$$E_{\min} = \frac{1}{N_1} \sum_{i=1}^{3} \mathrm{var}(G_{1i}) + \frac{1}{N_2} \sum_{i=1}^{3} \mathrm{var}(G_{2i}), \quad (5)$$

where $\mathrm{var}(G_{ji})$ is the variance of the ith *get3* vector component in the jth template region and N_j is the number of statistically independent *get3* vectors included in that region.

The right-hand side of Fig. 6 shows the result when this error limit is applied. Only one combination of parameters is below this error limit. There is a unique combination of experimental parameters which is a description for the experimental situation.

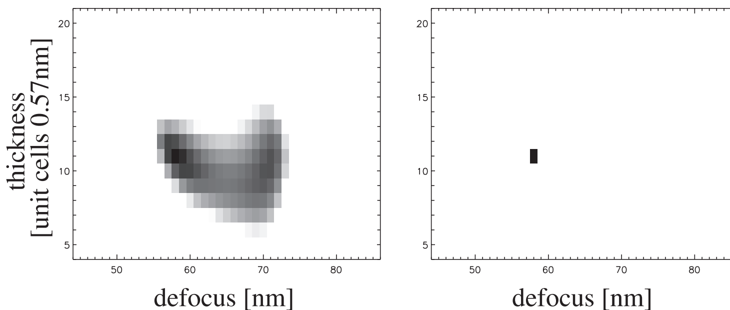

Fig. 6. Squared deviation of simulations from experimental values. Within the left figure, dark shading shows small deviation from experiment, while the right figure only shows deviations smaller than the error limit.

4. Neural networks

To maintain the relationship f between the *get3* vector and the composition parameter x (see Eq. (2)) we train a feed forward network with simulated examples (training set) of different experimental parameters (4). The used architecture is:

$$f(\vec{G}) = W_0 + \sum_{i=1}^{N_h} W_i \tanh\left(w_{i0} + \sum_{j=1}^{3} w_{ij} G_j \right), \tag{6}$$

where $\vec{W}$ and $\underline{\underline{w}}$ represent the network weights, i.e., the free parameters of the model. We use the RPROP algorithm [14] to do the network training (supervised batch learning).

4.1. Training set generation

The presented method is based on two different training sets. Both contain the linear adjusted (cf. Section 3.3) simulated *get3* vectors as input data. The first training set presents as output (supervised learning) the compositions of each involved simulation. The second training set contains as output the sample thicknesses. After the training, there are two neural networks, one for composition and one for thickness determination [3]. Because of their different influence on the *get3* vector composition variations can be distinguished from thickness variations throughout the image.

It is well known that for HREM images the major contribution of the noise in the image is due to an amorphous covering of the object. This covering results from the HREM sample preparation (ion milling). The random variation in the mass thickness of the covering leads to a random variation in the phase of the electron wave. Due to the lens aberrations, the imaging process does a spatial frequency filtering which leads to correlation in the noise throughout the image.

We simulate this amorphous object applying the random density object approximation (for description and comparison to other simulation models see [10]). Fig. 7 shows on the left-hand side a simulated AlGaAs interface structure and on the right-hand side the same simulation including 3 nm of amorphous object covering. The image distortion caused by the amorphous material is clearly to be seen.

[3] Note that the thickness in Section 3.4 is the average over the template regions and therefore has much less spatial resolution.

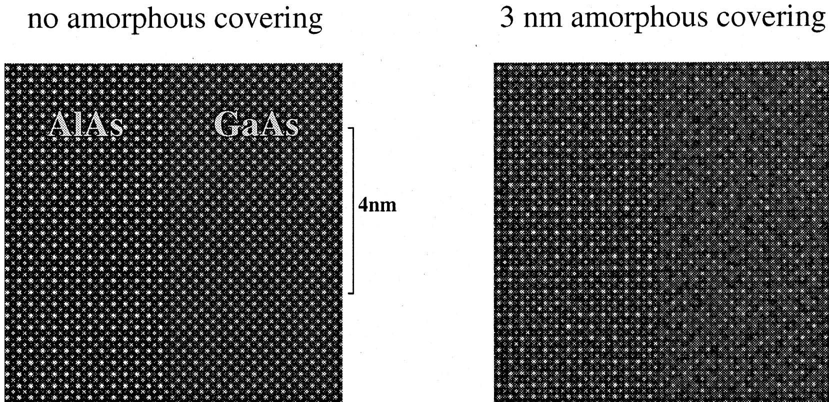

Fig. 7. Simulated images of AlGaAs interface. Right figure includes 3 nm of amorphous object covering.

4.2. Network architecture

The used architecture is fixed except for the number of hidden units N_h. This is a task for model selection algorithms. For both thickness and composition determination, the training time can be neglected compared to the time for simulations. Hence, we use stable but time-consuming test set validation for architecture selection (for more elaborate methods see [15]).

We tested architectures $N_h = 2$–30 and for each architecture size we trained 10 neural networks. Among all the resulting networks we select the one with best performance on a validation set (validation set patterns are excluded from training). It turned out that architecture sizes greater than $N_h = 30$ did not lead to benefit in error on a test set.

4.3. Comparison to classical methods

We compare the neural network-based method to classical methods of noise suppression. We tested all the methods with the same test scenario. The methods to be compared have to determine composition and thickness of simulated $Al_{1-x}Ga_xAs$ samples. Among the test samples composition x varies on the whole range [0, 1] in 10 steps. Sample thickness was from 9 to 15 unit cells (5.1–8.6) in steps of one unit cell (0.57 nm).The experimental defocus in the imaging process has the typical value of 58 nm. All the samples carried amorphous object covering with a thickness of 3 nm. These chosen parameter ranges are of high practical interest and chosen from parameters of experimental evaluations. Relative error with thickness was calculated in relation to the average thickness of 12 unit cells.

4.3.1. Method: minimum distance to simulated patterns

Similar to our estimation of average template parameters (Section 3.4) we simulate samples on a wide parameter range and seek the adjusted simulation with minimum distance to a local experimental *get3* vector. In contrast to the mentioned processing of lateral averaged *get3* vectors (Section 3.4), now methods are much more confronted with noise. The noise caused by the amorphous object covering leads to 16.2% error in composition determination and to 14.7% error in determination of sample thickness.

4.3.2. Method: projection perpendicular to principal component

We need to take benefit of the noise correlation. If we assume perfect correlation of noise then perfect noise discrimination is a projection orthogonal to the first principal component of noise (for principal component analysis PCA see [16]). For a detailed investigation of noise we calculated the PCA on *get3* vectors of a GaAs sample (thickness: 12 unit cells). The Eigenvalues w and Eigenvectors v of the correlation matrix were:

$$\begin{aligned} w_1 &= 0.00851 \quad \vec{v_1} = (-0.3817, 0.9229, 0.0509), \\ w_2 &= 0.00300 \quad \vec{v_2} = (0.9124, 0.3674, 0.1806), \\ w_3 &= 0.00045 \quad \vec{v_3} = (-0.148, -0.1154, 0.9822). \end{aligned}$$

The dominant Eigenvalue w_1 indicates a main direction in variation of noise. The corresponding Eigenvector $\vec{v_1}$ indicates an anticorrelation of the first two components of the *get3* vectors.

We calculate the plane perpendicular to $\vec{v_1}$ and project experimental *get3* vectors and adjusted simulated ones onto that plane. After that we search for the minimum distance simulation. The errors resulting from this method were 41.7% for composition determination and 15.1% for thickness determination. The increase of errors is due to the non-vanishing error variation in direction $\vec{v_2}$, which is not discriminated in contrast to the desired signal.

4.3.3. Method: optimized projection plane

A projection plane is numerically optimized with respect to performance on a validation set. For both composition and thickness determination an extra plane was adapted. Again, simulation and experiment are projected onto the plane and the simulation with minimum distance from experiment is selected. With this method the errors were 9.1% for composition and 5.7% for sample thickness.

4.3.4. Method: neural networks

As described in Section 4 we used a neural network-based method for determining compositions and thicknesses in our test set. The errors with the

Table 1
Errors of investigated methods for noise suppression

Noise suppression method	Error composition (%)	Error thickness (%)
Minimum distance to simulated patterns	16.2	14.7
Projection perpendicular to principal component	41.7	15.1
Optimized projection plane	9.1	5.7
Neural networks	6.7	2.5

neural network-based method were 6.7% for composition and 2.5% for sample thickness.

Table 1 shows the errors for the investigated noise suppression methods. The neural network-based method was of advantage in composition determination as well as in thickness determination. With regard to the thickness determination, the error with the neural network-based method was only half the error with the best classical method. The reason is that the neural network learns to suppress the error from the distorted training patterns. It takes benefit out of the correlation in noise.

5. Experimental results

Fig. 8 shows an AlGaAs interface structure. There are two template regions on both sides of the interface (see Fig. 4). With the parameter estimation

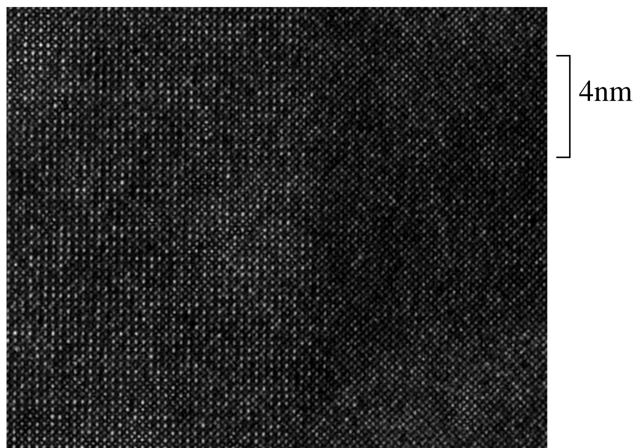

Fig. 8. Experimental HREM image of an AlGaAs interface structure.

method described in Section 3.4 the image showed the following experimental parameters:

composition of left template region:	$x_L = 0.65$,
composition of right template region:	$x_R = 1.0$,
thickness of left template region:	$t_L = 8.5$ nm,
thickness of right template region:	$t_R = 6.8$ nm,
defocus : $\Delta = 58$ nm.	

With a diffractogram analysis method (details see [10]) the thickness of the amorphous object covering was estimated 3.2 nm.

Fig. 9 shows the composition of the sample and the thickness of the crystalline part of the sample. Values are determined with spatial resolution of 0.28 nm. Within the graph the height of the columns indicates the local thickness and the greyscale quantifies the local composition of the sample.

The mean error for local composition determination was 5.9%. For the determination of the local thickness (crystalline) the mean error was 4.3% of mean thickness (8 nm). The composition in the ternary semiconductor (AlGaAs left-hand side of Fig. 8) varies strongly due to the stochastic occupation of one sublattice by two elements (random alloy fluctuations). The standard deviation of composition variation in the ternary alloy was in excellent agreement to a theoretical model.

Note, that the three-dimensional plot of Fig. 9 does not reflect the outer surface of the specimen. The determined thickness is only the thickness of the crystalline part. It does not include the amorphous object covering mentioned

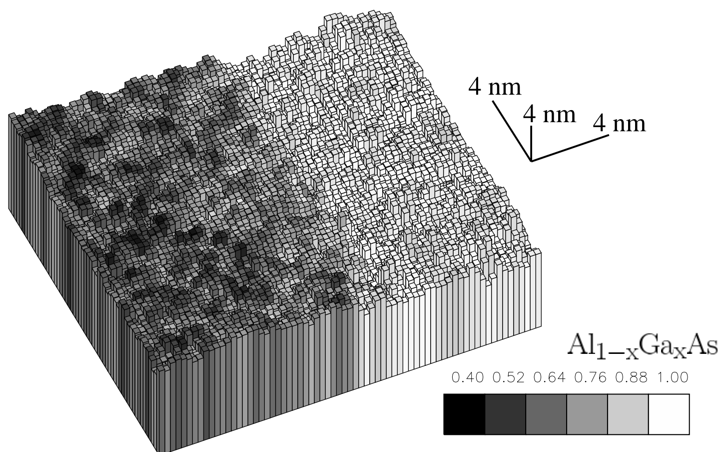

Fig. 9. Composition and thickness. Each column equals 1/4 of the unit cell area (0.28 nm^2). The height of the columns represent the sample thickness. The greyscale indicates the composition of the sample.

in Section 4.1. Moreover, the variation in thickness results from the roughnesses of the top *and* bottom surfaces of the sample and reflects neither of both individually.

The time consumption of the method on a PII 400 MHz was as follows: the simulation of training sets took up to one day. The training of the neural networks took 1 h (training set: 1080 examples) and the evaluation of the image only 30 s.

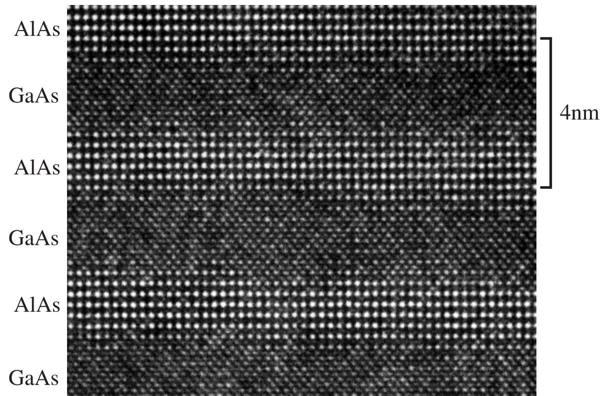

Fig. 10. Experimental HREM image of an AlGaAs Bragg-reflector.

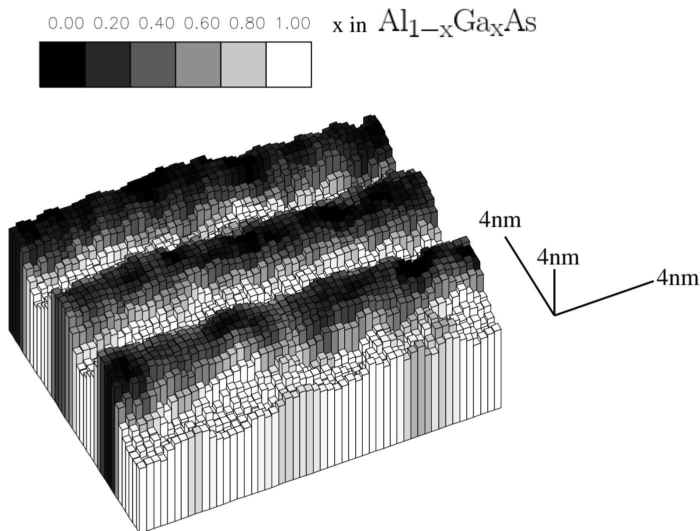

Fig. 11. Composition and thickness determination for the AlGaAs Bragg-reflector.

Fig. 10 shows a HREM image of an AlGaAs Bragg-reflector. Such heterostructures are used in laser diodes. The parameter estimation method (Section 3.4) indicated a defocus of 57 nm. The thickness of the amorphous object covering turned out to be 2.3 nm.

Fig. 11 shows the result of thickness and composition determination. Again each column equals to an area of 0.28×0.28 nm^2. The mean error for composition determination was 4.8% and for thickness determination 5.2% (of mean thickness 8.5 nm).

By intention this heterostructure does not differ significantly from a binary layer system (AlAs and GaAs). Within the thickness of the crystalline part of the sample there are striking differences between the AlAs and GaAs layers. This height difference is created during the HREM sample preparation applying ion milling.

6. Conclusion

The present paper describes a new neural network-based method of quantitative image processing in HREM. It renders the determination of local composition and thickness of compound semiconductor specimens. The stability with respect to the influence of amorphous object covering is an important criterion for methods that analyse microscope images. The suppression of this correlated distortion was carried out with several methods. It turned out that a neural network-based method was superior to classical methods. The application of neural networks led to a remarkable error reduction of up to 56%. The method has been applied to heterostructures of AlGaAs, which is illustrated by experimental examples.

Acknowledgements

We thank P. Werner for the HREM images and for critically reading the manuscript. This work has been generously supported by the Volkswagen–Stiftung under contract number I/71108.

References

[1] W. Kleber, H. Bautsch, J. Bohm, Einführung in die Kristallographie, Verlag Technik, Berlin, 1990.

[2] A. Ourmazd, F.H. Baumann, M. Bode, Y. Kim, Quantitative chemical lattice imaging: theory and practice, Ultramicroscopy 34 (1990) 237–255.

[3] D. Stenkamp, W. Jäger, Compositional and structural characterization of Si_xGe_{1-x} alloys and heterostructures by high-resolution transmission electron microscopy, Ultramicroscopy 50 (1993) 321–354.
[4] C. Kisielowski, P. Schwander, F.H. Baumann, M. Seibt, Y. Kim, A. Ourmazd, An approach to quantitative high resolution transmission electron microscopy of crystalline materials, Ultramicroscopy 58 (1995) 131–155.
[5] R. Hillebrand, Fuzzy logic approaches to the analysis of HREM images of III–V compounds, Journal of Microscopy 190 (1998) 61–72.
[6] R. Hillebrand, P.P. Wang, U. Gösele. Fuzzy logic applied to physics of III–V compounds, in: Proceedings of the Workshop on Breakthrough Opportunities for Fuzzy Logic, Tokyo, 1996, pp. 77–78.
[7] R. Hillebrand, P.P. Wang, U. Gösele, A fuzzy logic approach to edge detection in HREM images of III–V crystals, Information Sciences – Applications 93 (1996) 321–338.
[8] R. Hillebrand, P.P. Wang, U. Gösele, Fuzzy logic image processing applied to electron micrographs of semiconductors, in: P. Wang (ed.), Proceedings of the Third Joint Conference on Information Sciences'97, Duke University, Durham, I, 1997, pp. 55–57.
[9] H. Kirschner, R. Hillebrand, Neuronale Netze zur Kompositionsbestimmung von III–V Heterostrukturen in HREM Abbildungen, Optik (Suppl.) 1997, 74.
[10] H. Kirschner, HREM-Bildanalyse von III–V-Halbleiter-Schichtstrukturen durch quantitativen Bildvergleich experiment – simulation, Master Thesis, Martin-Luther-Universität Halle–Wittenberg, January 2000.
[11] P.A. Stadelmann, EMS – a software package for electron diffraction analysis and HREM image simulation in materials science, Ultramicroscopy 21 (1987) 131–146.
[12] P.A. Stadelmann, Image calculation techniques, Technical report, EPFL Lausanne, 1995.
[13] L. Reimer, Transmission Electron Microscopy, second ed., Springer, Berlin, 1989.
[14] M. Riedmiller, H. Braun, A direct adaptive method for faster backpropagation learning: the RPROP algorithm, in: H. Ruspini (Ed.), Proceedings of the IEEE International Conference on Neural Networks, San Francisco, 1993, pp. 586–591.
[15] H. Kirschner, Architekturabhängiges Lern- und Anpassungsverhalten bei Neuronalen Mehrschichtnetzen, Master Thesis, Institut für angewandte Physik der Universität Regensburg, 1997.
[16] I.T. Jollife, Principal Component Analysis, Springer, New York, 1986.