

# Optical track width measurements below 100 nm using artificial neural networks

R J Smith<sup>1</sup>, C W See<sup>1</sup>, M G Somekh<sup>1</sup>, A Yacoot<sup>2</sup> and E Choi<sup>1</sup>

<sup>1</sup> School of Electrical and Electronic Engineering, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

<sup>2</sup> National Physical Laboratory, Hampton Road, Teddington, Middlesex, TW11 0LW, UK

E-mail: [eejrjs@nottingham.ac.uk](mailto:eejrjs@nottingham.ac.uk)

Received 26 April 2005, in final form 20 September 2005

Published DD MMM 2005

Online at [stacks.iop.org/MST/16/1](http://stacks.iop.org/MST/16/1)

## Abstract

This paper discusses the feasibility of using artificial neural networks (ANNs), together with a high precision scanning optical profiler, to measure very fine track widths that are considerably below the conventional diffraction limit of a conventional optical microscope. The ANN is trained using optical profiles obtained from tracks of known widths, the network is then assessed by applying it to test profiles. The optical profiler is an ultra-stable common path scanning interferometer, which provides extremely precise surface measurements. Preliminary results, obtained with a 0.3 NA objective lens and a laser wavelength of 633 nm, show that the system is capable of measuring a 50 nm track width, with a standard deviation less than 4 nm.

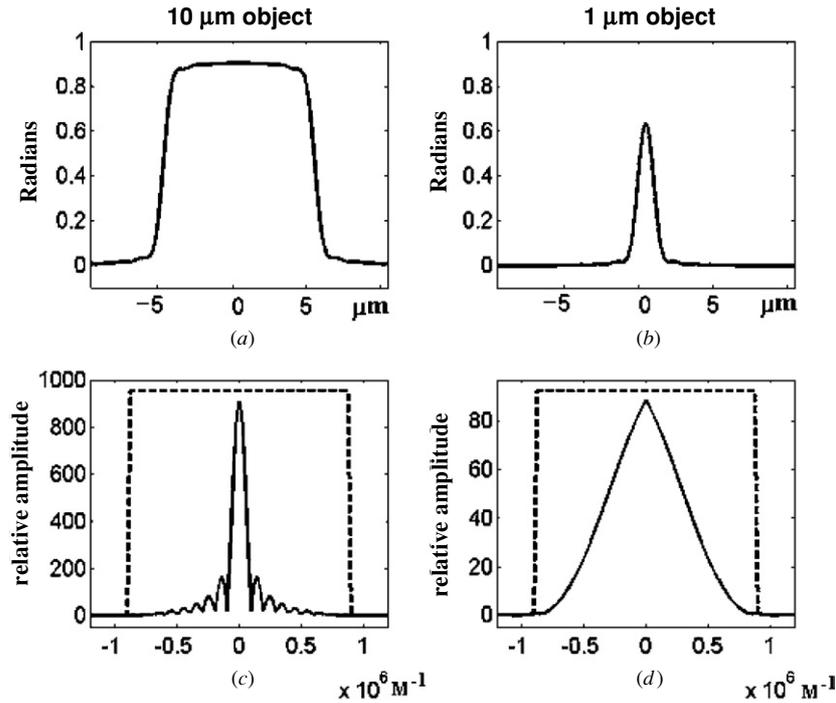
**Keywords:** optical track-width measurements, artificial neural networks, common path scanning interferometer

## 1. Introduction

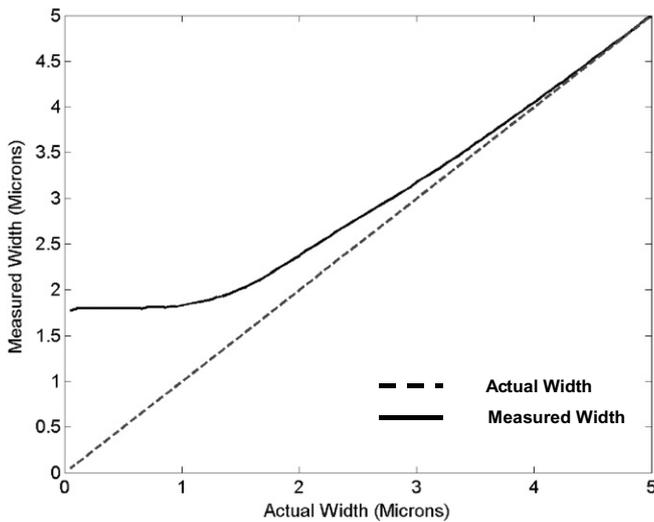
Calibrated line width standards are an important tool for many industries. Systems that are capable of providing reliable measurements are therefore useful in many application areas. One such area is in semiconductor industry, where the feature size has been decreasing over the years, and has reached a point that it is no longer possible to use traditional optical methods to obtain meaningful measurements. Other techniques seek to utilize shorter wavelength of radiation, including UV microscopy and scanning electron microscopy; or scanning near field techniques, such as atomic force microscopy, in order overcome this problem. These systems possess very fine lateral resolutions but they also suffer certain drawbacks: they are expensive compared to an optical microscope; they can be difficult to operate and particular care is needed to prevent the samples from being damaged. In addition, they are relatively slow and the data acquisition time can be long if a large sample area is to be interrogated. In this paper we will describe a technique, which combines an optical interferometric profiler with artificial neural networks, to provide track width measurements that are considerably smaller than the point spread function of the optical system.

It should be pointed out that, at the moment, the technique can only provide a single parameter, namely the track width, and it does not yield the surface profile of the sample. In its basic form, it can be used in conjunction with other measurement instruments to perform rapid measurements of samples. We are currently working to extend the application of the technique, with the aim of extracting other parameters relating to the shape of the samples.

The resolution of an optical system may be quantified by using the Rayleigh criterion, which gives the resolution to be half of the optical point spread function [1]. One can also consider an optical system as a spatially invariant linear system [2], and link its resolving power to the bandwidth of the system. Figure 1 is a simple illustration of this point. Figures 1(a) and (b) show the simulated responses of two tracks imaged with a partially coherent optical microscope, with an NA of 0.3 and an optical wavelength of 633 nm. The widths of the two tracks are 10  $\mu\text{m}$  and 1.2  $\mu\text{m}$  respectively. Figures 1(c) and (d) are the two corresponding image spectra, with the dotted line showing the pass band of the optical system. In the case of the 10  $\mu\text{m}$  track, most of the significant frequency components of the object is inside the bandwidth of the optical system. This is clearly not the case with the narrow



**Figure 1.** Simulated phase profiles and their spectra: (a) phase profile of a 10  $\mu\text{m}$  track, (b) phase profile of a 1  $\mu\text{m}$  track, (c) spectrum of 10  $\mu\text{m}$  track and (d) spectrum of 1  $\mu\text{m}$  track.



**Figure 2.** Simulated data: measured track response as a function of the actual track width.

track. Indeed, if we reduce the track width further, the image produced by the system would remain substantially similar to that of figure 1(b), and the object is considered beyond the resolution limit of the system.

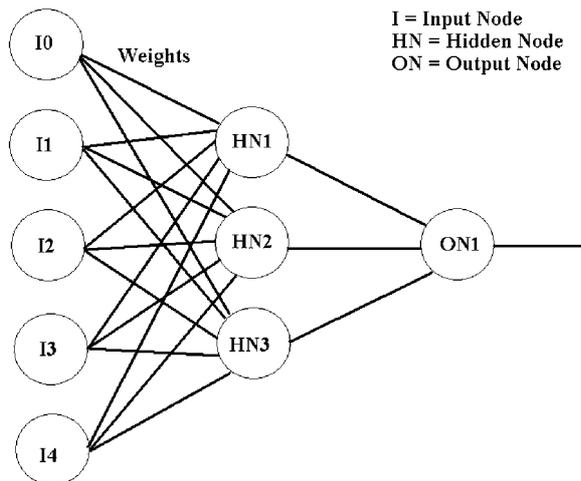
Figure 2 shows the simulated full width half maximum (FWHM) of tracks as a function of the input track width for the system described above, i.e. NA of 0.3 and wavelength of 633 nm. The dashed line represents the actual track width. It is apparent that, when the track width is below a certain value, the measurement would be much more sensitive to the presence of system noise and is no longer reliable.

There has been much research aimed at overcoming the diffraction limit, thus extending the measurement capability of optical instruments. Many of these techniques seek to

reconstruct the high frequency components of the object spectrum that are originally outside the bandwidth of the optical system [3, 4]. The basis of these techniques is the principle of analytic continuation [5], which stipulates that from the knowledge of an analytic function over a finite interval, the function over its entirety can be recovered [6]. It can be shown readily that the Fourier spectrum of a spatially bounded object is analytic [4], so that the object spectrum outside the system bandwidth can be reconstructed. The extent of the spectrum extension that can be achieved depends on the actual algorithm employed. It also depends critically on the accuracy and precision of the captured image. The resolution improvement that can be attained in practice is limited.

The problem of super-resolution has been studied using the approach of information content [7], and the total amount of information contained in a set of measured data is fixed. Post-measurement improvement of a particular aspect of the data, can only be achieved at the expense of other parameters. Hence, extending the spectrum of the measured data will result in a degradation, for example, of the signal-to-noise ratio of the signal. This can explain the limited capability of the various spectrum extension techniques. In contrast, since our aim is to extract one parameter only, we can expect to achieve a much greater improvement as will be demonstrated later.

In the next section, we will first state the general properties of ANNs. This will be followed by a detailed discussion of the networks constructed for this project. The optical interferometer used for extraction of object information has already been reported in an early publication [8]; so, only the salient features of the system will be described in section 3. The operating procedure of the system and experimental results will be presented in section 4. The last section will contain discussion of the technique and also the conclusions.



**Figure 3.** Layout of a simple feed forward artificial neural network.

## 2. The ANN approach

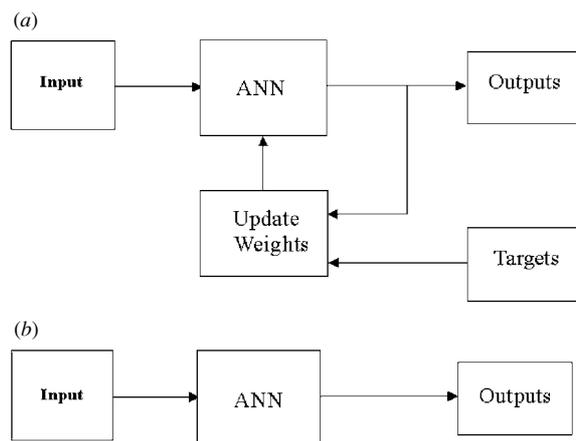
Artificial neural networks are biologically inspired computational units, built around a simple model of the neuron. They are arranged in groups to produce a network of many computational units. ANNs have been used to generate generalized models of input and target data in many applications. For example image compression [9], modelling complex process dynamics [10, 11], nonlinear time varying systems [12], to name but a few.

There are different types of ANNs, which are usually divided, according to the way in which they learn, into supervised and unsupervised networks. Each category is subdivided into sets of other networks that share similar properties, for supervised learning there are feed forward, feed back (recurrent) and competitive networks. Each have their own advantages and disadvantages depending on the application. Feed forward networks such as the multilayer perceptron [13] have the advantage of being easy to train and can approximate any nonlinear function given enough processing elements; feedback architectures are good for sequential behaviour problems, such as speech recognition etc, but can be time consuming and difficult to train due to the feedback loops.

The ANN used for our application is a feed forward network, which has been shown to be particularly suitable for extracting small changes of the input signal [14]. It has an input layer, a hidden layer and an output layer. This is illustrated in figure 3.

When in operation, a data pattern is presented to the input, the network feeds the values through the network from front to back, calculating the weighted sum of each input at each node. This weighted sum is the input to an activation function, which introduces nonlinearity into the network making the network a much more powerful computational unit. The outputs are fed to the next layer. At the output layer the value obtained, in our case, is the track width of the presented input pattern.

The values of the weights are determined during the training stage. This is achieved by presenting inputs and their *known* outputs to the network. The weights are updated by an iterative, least square error process known as back propagation. The error at the output of the network is



**Figure 4.** Schematic of ANN in (a) training and (b) operation.

propagated backward through the system to the input, at each layer, the weights are adjusted to reduce the overall error at the output. This process is repeated until a predefined condition is met. The way in which the weights are actually updated depends on the training rule used. Both the feed forward network and the back propagation algorithm are discussed extensively in the literature, some can be found in the following references [13, 15]. The operation of the ANN in both training mode and normal use are shown in figure 4.

The number of training patterns and selection of network size are important considerations. There are no hard and fast rules for selecting these parameters due to the number of factors that can influence network design e.g. training patterns available, complexity of model, type of ANN, activation function used, etc. In our case the number of inputs will depend on the data type and will be experimentally determined. As a general rule of thumb 30 times as many patterns as weights in the network are required to reduce the chance of over fitting.

Over fitting, which is also known as over training, is usually caused by having too few input patterns or an overly complex network with noisy data. A properly trained network produces a generalized model of the input and target relationship. This means that it will produce valid answers for any pattern presented within its training range even if it was not contained in the training set. If a network has been over trained, then a pattern that was not contained in the training set will produce larger errors when presented to the network, as the network has memorized the training patterns instead of learning a general rule.

One method to avoid over fitting is to have a large data set, having 30 times more patterns than weights in the network should reduce the chance of over fitting, although this is not always possible due to limited data sets. Other approaches can be used to overcome this problem when the number of data sets are a problem, such as correct model selection, early stopping, Bayesian learning, weight decay and jittering (training with additional noise) [16, 17]. It is difficult to quantify the number of patterns required using these methods; however, by adopting the early stopping and jittering approaches, we have trained successfully with as few as 15 patterns.

The networks used for these measurements have typically contained eight input values, five hidden nodes and one output node. The node activation functions are tansigmoid functions

(hyperbolic tangent in this case) [15] and the training rule that was used is based on the Levenberg Marquardt optimization rule [18]. This training rule provides faster training than other methods such as simple gradient descent methods, without compromising the reliability of the system. The network layout was chosen because it provided repeatable training and good results for our data, although other network configurations may also perform well.

In our system, the patterns used for training the ANN are split into two sets. One set, normally 75% of the total, is used for training the ANN. The remaining is used during training as a validation set. As there are relatively few distinct input patterns, due to the number of samples available, an early stopping technique is employed. During the training process the error in the validation set is monitored; if the trend of the error increases the training is stopped, as an increase in the error of the validation set implies that the network has started to memorize the input/output patterns and lose generalization.

The format of the input data can critically affect the performance of the ANN, and is application dependent. The data format that is suitable for our system will be discussed in section 5. In the next section, the optical system used for the extraction of the object data will be presented.

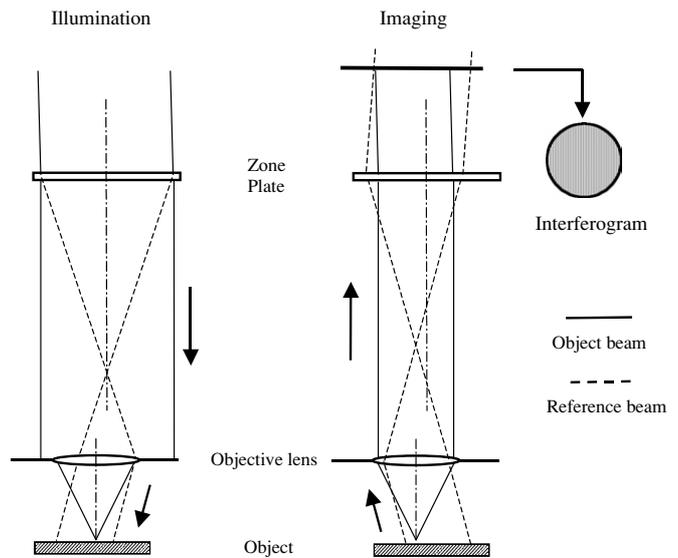
### 3. The optical system

In order to ensure successful training of the ANN, the optical system is required to produce profiles that are stable, of high signal-to-noise ratio, and repeatable both in the short and long term. A scanning interferometer is preferred since many samples of interest are phase objects and do not present variations in reflected intensity.

The system used to obtain surface profiles is an ultra stable common path scanning optical interferometer. Because of the common path nature of system, effects of microphonics due to vibrations and thermal gradients are greatly reduced, thus allowing the system to perform close to the shot noise limit.

The system uses a computer generated holographic (CGH) diffractive element as the beamsplitter. The arrangement between the objective lens and the hologram is shown in figure 5. The CGH creates two output beams from a collimated input beam. The first is an unaltered zero order which is focused onto the sample by the objective, and acts as the sample probe beam. The second is a first-order beam, converging to the back focal plane of the objective. The objective then collimates the beam onto the sample surface at some angle depending on the lateral offset of the hologram with respect to the optical axis, and this beam serves as the reference beam.

The two returning beams are recombined by the hologram and interfere to form straight fringes, the frequency of which is set by the angle of incidence of the collimated beam at the sample surface. Local surface height variations will change the phase of the probe beam, whereas the average phase of the reference will remain essentially unchanged. To extract the surface information, the interferogram is recorded using a CCD camera. Fourier transformation is then applied to the interferogram. It can be shown readily that the phase and amplitude of the sample are given by the Fourier component due to the fringe frequency, and the profile of the surface



**Figure 5.** Light paths of the ultra-stable scanning optical interferometer.

is obtained by scanning the sample. It should be noted that the two light beams traverse the optical system through similar paths, and the effects of microphonics will largely be cancelled when the two beams interfere. This will improve the stability of the system, and allows the system to perform close to its fundamental limits. Both are critical considerations for the application of this project. Figure 6 shows the image of the fringes recorded by a CCD camera. The image was taken using a wavelength of 633 nm and an objective of 0.3 NA.

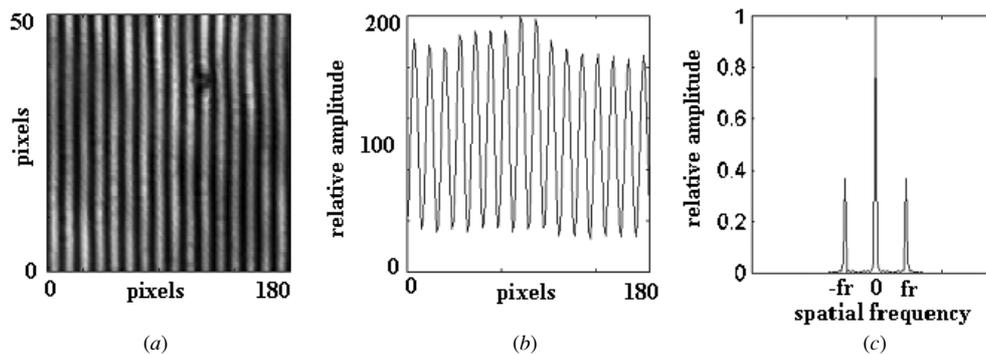
An example of the use of the system is shown in figure 7, where it was used to measure a 17 nm high and 40  $\mu\text{m}$  pitch phase grating. The phase profile was obtained by scanning 100  $\mu\text{m}$  across the grating, and the high signal-to-noise ratio for the system is clearly demonstrated. The signal-to-noise ratio for the amplitude signal is around 3000:1 and the phase noise has a standard deviation of around 0.5 mrad, both were obtained in standard laboratory conditions.

## 4. Experimental results

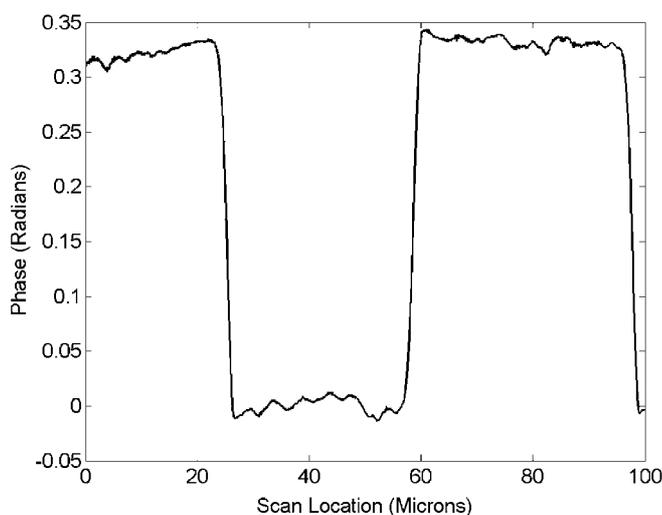
### 4.1. BCR standard sample

The common path interferometer described above was used to measure a line width standard, known as the BCR sample, provided by National Physical Laboratory (NPL), UK. The sample consisted of a glass substrate coated with a 100 nm thick chromium. Thirteen tracks of widths from 0.273  $\mu\text{m}$  to 2.18  $\mu\text{m}$  and length 20  $\mu\text{m}$  were etched through the metal layer. A thin layer of aluminium was evaporated to cover the structure, thus making it a purely phase object. Each track was scanned 6 times to build up a set of 78 phase profiles that could be used for training and testing the ANN. It should be noted that the profiles were selected randomly into the training and testing sets, and therefore they might not cover the track width range evenly. This random division process was adopted to further ensure proper training of the networks.

Figures 8(a) and (b) show the phase profiles of two tracks, with widths of 2.18 and 0.273  $\mu\text{m}$ . Figures 8(c) and (d) show the corresponding phase spectra, where the dotted line is the



**Figure 6.** Outputs from optical interferometer: (a) typical fringe pattern captured using the CCD camera, (b) line trace across the fringe pattern and (c) Fourier spectrum of the fringe pattern. Object information is derived from the amplitude and phase of the sideband at frequency  $fr$ .



**Figure 7.** Profile of a phase grating obtained with the scanning optical interferometer, height = 17 nm and pitch = 40  $\mu\text{m}$ .

pass band of the optical system. Throughout the experiment, the wavelength used was 633 nm and the objective had a numerical aperture (NA) of 0.3, giving a system resolution of almost 1.3  $\mu\text{m}$ . Each profile contained 500 scan points with a scan increment of 40 nm

It is important that the signals are in an appropriate format before they are used as the ANN inputs. The exact format depends on the actual application. For this project, different data formats have been tried, and the one described below yielded the best results. The data processing procedure is:

- The data are normalized, using one of a number of methods, which will be discussed below;
- The data are shifted so that the track is centred with respect to the window;
- A Hann window [19] is applied to the data, in order to minimize errors caused by edge effects;
- The function is zero padded to an appropriate length;
- The profile is differentiated before a discrete Fourier transform is applied;
- The amplitude spectrum of the differentiated profile is then sampled at eight equally spaced locations inside the system bandwidth. The sampled values are then used as the input of the ANN.

The critical steps of the above process are the last two in the list. Different data formats have been tried as the network input, including the profiles, the normalized profiles, the differentials of the profiles, the spectra of the profiles and the spectra of the differential profiles. The last one yields by far the best results. This is similar to our experience when we used ANNs to measure thermally induced micro-deformation [14]. Again, using the spectrum of the differentiated data as the network input produced the best results. As can be expected from figure 1, it is necessary to represent strongly the high frequency components, which will enhance the differences in width between the narrow structures.

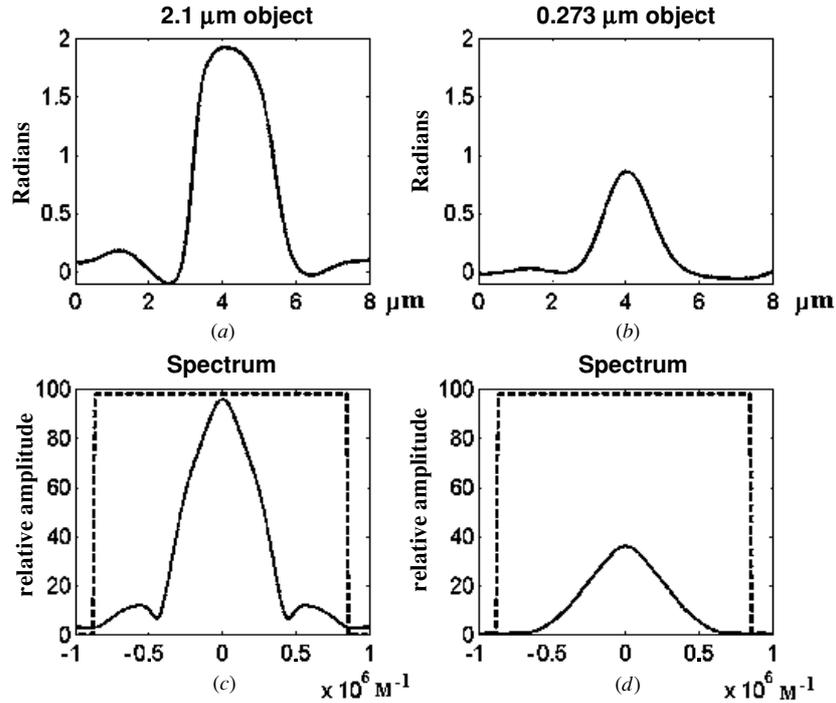
The other important factor is the number of input nodes. It is found experimentally that the optimum number of input nodes is around 8. If too few are used, the network could not be trained successfully. If too many, the network would attempt to memorize, rather than to generalize the measurement situation.

The whole input data set is then scaled so that the maximum value of all of the input points is 0.8, as this helps the stability of the training. The same scaling was applied to the target widths because the targets need to be in the range of the activation function, which in this case is  $\pm 1$ . Using a value of 1 for the scaled level can cause the network to be unstable since in order for this value to be attained the input into the activation function has to be infinity.

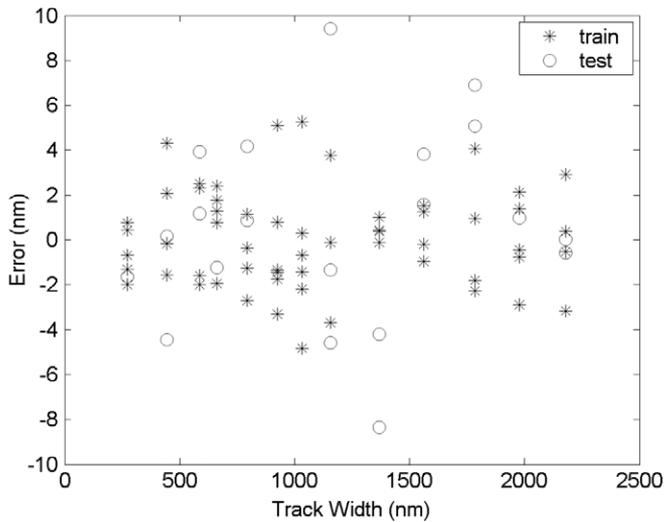
The target widths used in the training process were the calibrated values obtained from measurements made at NPL using the OPTIMM system for line width measurement [20].

The results obtained for this sample are shown in figure 9. Each point on the graph represents the difference between the track widths as measured by OPTIMM and the value provided by the trained ANN. The crosses represent the training set and the circles the validation set. The mean and the standard deviation of the differences between the measured and the true values are tabulated and shown in table 1. Also shown are the maximum and minimum differences. The five rows correspond to five different types of normalization procedures, which are

- (1) No normalization: the object phase profiles are processed as described, and the values of the differential spectrum are used directly as the ANN inputs;
- (2) Normalized with respect to the peak of the phase profile: before the processing of the data, the phase profiles are



**Figure 8.** Phase profiles of the BCR sample and their spectra, obtained using the scanning optical interferometer: (a) profile, 2.18  $\mu\text{m}$  track width, (b) profile, 0.273  $\mu\text{m}$  track width, (c) spectrum of (a), and (d) spectrum of (b). Track height = 120 nm.



**Figure 9.** ANN results: BCR sample, error in standard deviation for both the training and testing sets.

normalized so that individual profiles would have peak values of unity;

- (3) Normalized with respect to the sum of the absolute values of the phase profile;
- (4) Normalized with respect to the sum of the square of the phase profile;
- (5) Normalized with respect to the sum of the differential spectrum.

From table 1, it can be seen that, in terms of standard deviation, normalization type 1 and 4 produces the best results. These results, together with the fact that eight input nodes is the optimum number, indicate that the ANN is relying on both the shape of the phase profile, as well as the total

amount of ‘energy’ each profile contained, to provide accurate measurements, although more work is required in this area. The results presented in this paper are all generated using the raw phase profiles, with no normalization applied to the data.

Figure 9 shows that the errors are spread evenly over the entire track width range, moreover the errors in the test data while worst than the training data are not substantially inferior. The lateral resolution of the optical system, according to the Rayleigh criterion, is 1.3  $\mu\text{m}$ . This compares to the narrowest track width of 0.273  $\mu\text{m}$ . A measurement precision of less than 5 nm standard deviation demonstrates the capability of the technique.

#### 4.2. Measurement of silicon sample

A second sample of purely silicon tracks, with widths from 60 nm to 480 nm in steps of 20 nm and height of 45 nm was also measured. The configuration of the optical system was kept unchanged from the configuration used for the BCR sample, with an optical wavelength of 633 nm and an objective NA of 0.3. The normalization procedures used are the same as for the BCR sample discussed above. The results of the training are shown in table 2, and figure 10 shows the spread of the error.

Once again, normalization type 1 and 4 produce the best results. However, the magnitudes of the errors are even smaller than the other set. This may be attributable to the cleanliness of the sample. The narrowest track on this sample has a width of 60 nm, which is some 22 times smaller than the resolution limit of the system! It should be pointed out that 60 nm is not the limit of the capability of the technique, but rather, it is the narrowest track available to us.

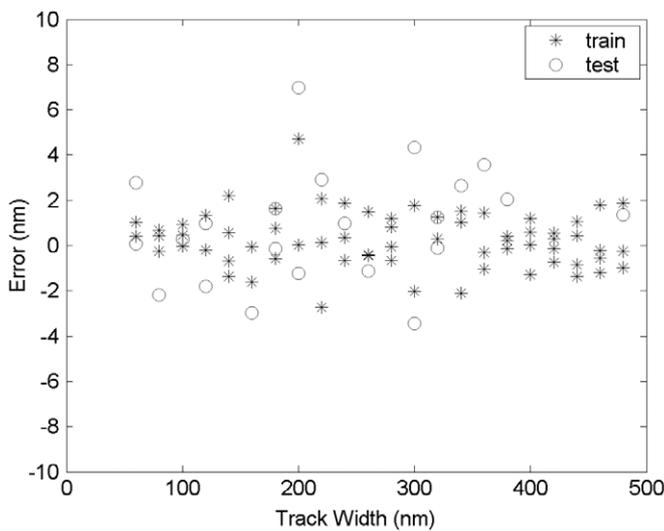
For both samples, the amplitude profiles of the tracks were also used to train ANNs. The resulting ANNs produced

**Table 1.** Results for BCR sample, all values in nm.

Norm type	Training		Validation		Peak error	
	Std error	Mean error	Std error	Mean error	Min	Max
1	2.19	0.03	4.25	0.59	-5.92	6.67
2	9.53	5.62	10.86	0.16	-17.94	20.51
3	8.57	-1.23	16.62	2.03	-23.85	26.47
4	6.18	2.56	3.28	2.15	-8.09	15.09
5	6.23	-1.39	14.55	-0.12	-21.14	22.23

**Table 2.** Results for silicon sample, all values in nm.

Norm type	Training		Validation		Peak error	
	Std error	Mean error	Std error	Mean error	Min	Max
1	1.22	0.25	2.57	0.69	-3.13	4.91
2	10.03	-3.79	14.49	-1.48	-25.33	22.89
3	11.67	0.54	16.38	-3.19	-30.90	25.80
4	1.83	-0.14	5.95	2.76	-4.37	13.62
5	19.08	9.63	23.25	11.02	-37.93	55.47

**Figure 10.** ANN results: Si, error in standard deviation for both training and testing sets.

errors several times greater than those obtained from the phase profiles. This is hardly unexpected as the amplitude profiles were merely due to scattering at the edges of the shallow tracks.

#### 4.3. Tolerance of the ANNs

In order to investigate the tolerance of the ANNs to the input data, the training of the ANNs has been altered. With the two examples shown above, the widths of the tracks in the training were nominally the same as those in the testing sets. Other networks have been trained, with up to three different tracks omitted from the training. When the networks were applied to the complete testing sets, results similar to those shown in figures 9 and 10 were obtained. The one exception was when the omitted tracks were at the ends of the range, where much larger errors were resulted. This is essentially a reiteration that ANNs are very good at interpolation but generally rather poor at extrapolation.

## 5. Discussion and conclusions

In this paper, we have described the use of an ultra-stable scanning interferometer and artificial neural networks for accurate and precise measurements of track widths that are substantially smaller than the resolution limit of the optical system. With a 0.3 NA objective lens and a laser wavelength of  $0.633 \mu\text{m}$ , track widths in the region of 50 nm can be measured, with an uncertainty less than 3 nm. It should be noted that this value includes error associated with the technique, variations in the line widths of the samples, and the precision of the scanning stage. The stage used in the optical system was the x-y piezo flexure nanopositioner with capacitive sensors (P-731.20), with a stated resolution of less than 1 nm. Work is underway to incorporate an interferometer to the system. By monitoring the movement of the stage with the interferometer, the measurement precision should be further improved.

In order to train a network successfully, the input data must be of very high quality. This demands the optical system to produce repeatable measurements that are of high signal-to-noise ratio. Departure from these conditions will result in a spread of the error, and in some cases, may cause the training to fail. The ultra-stable scanning interferometer employed in the system meets these demands. The common path nature of the interferometer renders the system insensitive to microphonics and low frequency changes in the environment. Indeed, the results shown above were obtained with the system located in a general laboratory, without any specific measures to isolate the system from the surroundings. By operating the system in a more controlled environment, together with using a higher NA objective lens and shorter laser wavelength, we are confident that the performance of the system can be improved further. Accurate measurement of track widths in the region of 10 nm is a distinct possibility.

In section 4, we have discussed the effects of the various normalization methods on the operation of the ANNs. It has been found that, for pure phase tracks, the spectra of the differential profiles are most suitable as the network inputs, and no normalization is required. We have trained networks with different input patterns. These include reducing the number of input nodes, and also using only a portion, rather than the

whole spectrum as the inputs. It is found that the scheme used for the experiments above, eight points spread evenly over the spectrum, provided the most precise and reliable results. This suggests that the networks require both the height information contained in the profiles, as well as changes in the relative amplitudes of the high spatial frequency components.

As mentioned before, the technique only provides one single object parameter. It relies on the existence of a set of known samples, which may be obtained with the help of, for example, an AFM. Since AFM data would be used to train the neural networks, the system described here could be thought of as a low cost virtual AFM. Consequently, it is envisaged that, in its current form, the system will provide rapid and precise track width measurements and will be ideal as a quality assurance tool.

We are currently working to increase its application areas of the system. Results obtained from simulations show that the technique can equally be applied to tracks that are close to one another, and also non-rectangular tracks. In some cases, more than one network may be required to fulfil the measurement. Our aim is to create a group of networks, which will classify the objects, according to their types and ranges, and will direct the data to the most appropriate network for proper measurement.

Another area of work concerns the rigorous testing of the optical system and the ANNs, in order to be certain that the technique can be used regularly for calibrating linewidth standards. The allowed deviations between the master standard and the sample under evaluation need to be quantified, and an uncertainty budget needs to be written. Work is in progress to address these issues.

## Acknowledgments

RS thanks the EPSRC and the National Physical Laboratory for a studentship. Part of this work was carried out under the DTI Programme for Length Measurement 2002–2005.

## References

- [1] Born M and Wolf E 1999 *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* 7th edn (Cambridge: Cambridge University Press)
- [2] Goodman J W 1996 *Introduction to Fourier Optics* 2nd edn (New York: McGraw-Hill)
- [3] Toraldo Di Francia G 1955 Resolving power and information *J. Opt. Soc. Am.* **45** 497–501
- [4] Harris J L 1964 Diffraction and resolving power *J. Opt. Soc. Am.* **54** 931–6
- [5] Whittaker E T 1902 *A Course of Modern Analysis: An Introduction to The General Theory of Infinite Processes and of Analytical Functions* (Cambridge: Cambridge University Press)
- [6] Morse P M and Feshbach H 1953 *Methods of Theoretical Physics part I* (New York: McGraw-Hill)
- [7] Cox I J and Sheppard C J R 1986 Information capacity and resolution in an optical system *J. Opt. Soc. Am. A* **3** 1152–8
- [8] Sawyer N B E, See C W, Clark M, Somekh M G and Goh J Y L 1998 Ultrastable absolute-phase common-path optical profiler based on computer-generated holography *Appl. Opt.* **37** 6716–20
- [9] Ma L and Khorasani K 2002 Application of adaptive constructive neural networks to image compression *IEEE Trans. Neural Netw.* **13** 1112–26
- [10] Parlos A G, Chong K T and Atiya A F 1994 Application of the recurrent multilayer perceptron in modeling complex process dynamics *IEEE Trans. Neural Netw.* **5** 255–66
- [11] Faller W E and Schreck S J 1995 Real-time prediction of unsteady aerodynamics: application for aircraft control and manoeuvrability enhancement *IEEE Trans. Neural Netw.* **6** 1461–8
- [12] Levin E 1993 Hidden control neural architecture modeling of nonlinear time varying systems and its applications *IEEE Trans. Neural Netw.* **4** 109–16
- [13] Swingler K 1996 *Applying Neural Networks—A Practical Guide* (San Francisco: Morgan Kaufman Publishers)
- [14] Pitter M C, See C W and Somekh M G 2001 Subpixel microscopic deformation analysis using correlation and artificial neural networks *Opt. Express* **8** 322–7
- [15] Haykin S 1999 *Neural Networks—A Comprehensive Foundation* 2nd edn (Englewood Cliffs, NJ: Prentice-Hall)
- [16] Sarle W S 1995 Stopped training and other remedies for overfitting *Proc. 27th Symp.: Interface of Computing Science and Statistics* pp 352–60  
<ftp://ftp.sas.com/pub/neural/inter95.ps.z>
- [17] Neal R M 1996 *Bayesian Learning for Neural Networks* (New York: Springer)
- [18] Hagan M T and Menhaj M 1994 Training feedforward networks with the Marquardt algorithm *IEEE Trans. Neural Netw.* **5** 989–93
- [19] Rabiner L R and Gold B 1975 *Theory and Application of Digital Signal Processing* (Englewood Cliffs, NJ: Prentice-Hall)
- [20] Nunn J, Mirande W, Jacobsen H and Talene N 1997 Challenges in the calibration of a photomask linewidth standard developed for the European Commission *VDE-VDI Conf. Proc.: Mask Technology for Integrated Circuits and Micro-components* pp 53–68