

Line Width Measurement below 60nm Using an Optical Interferometer and Artificial Neural Network

Chung W See^{*1}, Richard J Smith^{*}, Michael G Somekh^{*}, and Andrew Yacoot⁺

^{*}University of Nottingham, School of Electrical and Electronic Engineering, University Park, Nottingham, NG7 2RD, UK.

⁺Division of Engineering and Process Control. National Physical Laboratory, Queens Road, Teddington, Middlesex, UK, TW11 0LW

ABSTRACT

We have recently described a technique for optical line-width measurements. The system currently is capable of measuring line-width down to 60 nm with a precision of 2 nm, and potentially should be able to measure down to 10nm. The system consists of an ultra-stable interferometer and artificial neural networks (ANNs). The former is used to generate optical profiles which are input to the ANNs. The outputs of the ANNs are the desired sample parameters. Different types of samples have been tested with equally impressive results. In this paper we will discuss the factors that are essential to extend the application of the technique. Two of the factors are signal conditioning and sample classification. Methods, including principal component analysis, that are capable of performing these tasks will be considered.

Keywords: Artificial Neural Network, interferometer, line-width measurement, principal component analysis

1. INTRODUCTION

Dimensional measurement has long been an important application for optical techniques. Systems based on fringe projection and optical triangulation have been used extensively for measuring the shapes and sizes of objects. At the other end of the application spectrum is critical dimension (CD) measurement, where sub-micrometre measurement is required. For dimensional metrology, optical techniques possess many attractions: they are non-contact, reliable and relatively easy to operate. Their measurements can be related to international recognised standards, and are traceable when monochromatic light source is used. However, their applicability for CD measurement has diminished greatly in recent years. Because of the rapid advance in modern industry, the dimensions of many features of interest are so small that they

¹ chung.see@nottingham.ac.uk, Tel +44 (0) 115 9515535, fax +44 (0) 115 951 5616, www.eee.nottingham.ac.uk

are beyond the resolution limit of optical techniques. To this end, the use of scanning electron microscopy and atomic force microscopy is becoming more widespread for CD metrology.

Over the years, there has been much research aimed at increasing the lateral resolution of optical systems. Many of the proposed methods attempt to extract object information from the image that was originally outside the optical system bandwidth. The theoretical background behind the techniques was established in a number of publications. [1,2,3]. Different implementations have been proposed to realise the resolution improvement, some based on analytical methods [3,4] and others are iterative [5]. Naturally, the performance of all these methods is greatly affected by system noise. By using the prolate spheroidal wave functions [6,7], Rushford and Harris [8] examined the influence of various types of noise on the improvement achievable. Another approach investigated the problem from the view point of information theory [9,10], and introduced the concept that the information content of an image is a fixed quantity. Processing of the image may increase the effective bandwidth of the system, but it can only do so at the expense of other parameters, such as a deterioration in the SNR of the data. In practice, the resolution improvement that can be achieved is limited.

We tackle this problem from a different angle. Instead of attempting to increase the effective bandwidth of the system, we aim to measure the dimensions of the sample directly. We use an ultra-stable optical interferometer [11] to produce samples profiles, which are then input to suitably trained artificial neural networks to produce the sample dimensions. In a previous publication [12] we have shown measurements of track widths down to 60 nm, with an repeatability of 2 nm. The optical system used had a numerical aperture of 0.3, and a HeNe was used as the source. The point spread function (psf) of the system was 2.6 μm in diameter ($\text{NA} = 0.3$, $\lambda = 633 \text{ nm}$), which was about 43 times the width of the track measured.

The objective of this paper is to investigate the possibility of extending the application of the technique. We will study the critical factors that need satisfying if the system is to measure multiple sample parameters, such as width, separation, depth and local gradient. In the next section, we will summarise the salient points of our technique. In section 3 we will use results obtained before to demonstrate the capability of the system. We will then discuss issues relating to the measurement precision and noise, and the classification of different types of samples.

2. OPTICAL INTERFEROMETER AND ANN

The system principles have been reported previously [12] and we will only describe the main points here. The system consists of a scanning optical interferometer and artificial neural networks. The requirement of the interferometer is to produce highly repeatable surface profiles with high signal to noise ratio. The profiles are input to the ANNs that have been trained for the measurements.

The optical system is a common path scanning interferometer, which we developed previously [11]. It uses a computer generated hologram as the beam-splitter. This allows the two interfering beams to traverse similar path through the system and reduces greatly the effects of vibration. The interferogram thus formed consists of a set of parallel fringes, and the surface information is contained in the phase and contrast of the fringes. The fringe pattern is captured with a CCD camera and processed in a PC. Because of the common path nature of the system, the interferometer is extremely stable, with the long term drift (over several hours) and the short term phase noise to be 1 mrad and 0.01 mrad respectively. The sample

stage is a piezo flexure nanoscanner (PI-P517) with capacitor position sensors, and has a 1 nm scan resolution. By scanning the sample in the x-y plane, profiles of the samples are built up.

Before the ANNs can be applied to the measured profiles, they need to be trained. This will require a set of reference samples of known dimensions, covering the measurement range of interest. The training procedure is illustrated in figure 1. The reference samples, typically 15 or so, are measured using the optical system. The signal conditioner firstly differentiates the profiles and then calculates their spectra, which are used as the inputs to the ANN. The weights of the network are initially assigned with small random values. The ANN's output for each profile is compared to the target values, which are the known dimensions of the samples. The error signal e is feedback to the network, and the weights are adjusted using the back propagation algorithm [13], to reduce the error. This process is repeated, using the same set of samples, until some pre-defined conditions are met and the process is then terminated. The resulting ANN is then trained for that particular measurement purpose. For different applications, separate ANNs will need to be trained.

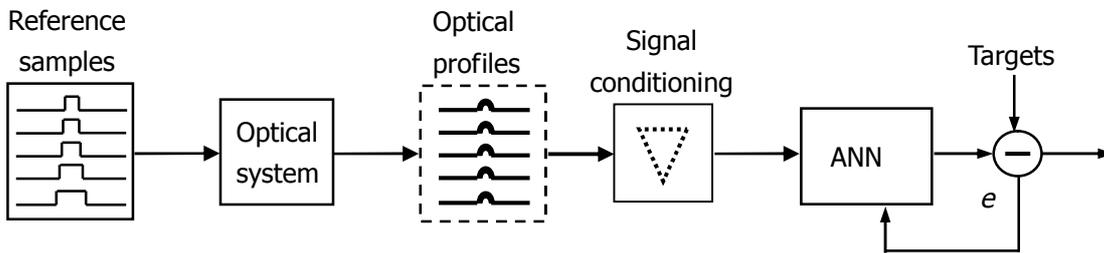


Figure 1. Schematic showing the training of ANN

The type of ANN used in our system is a simple multiple layer feed-forward network, with eight input nodes and five hidden nodes. The number of outputs depends on the required measurement and is either one or two. Figure 2 is a schematic of the ANN.

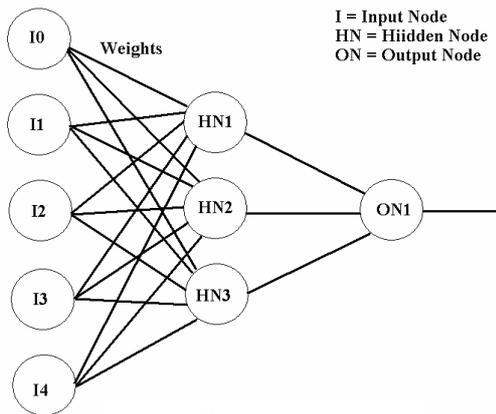


Figure 2. Schematic of artificial neural network

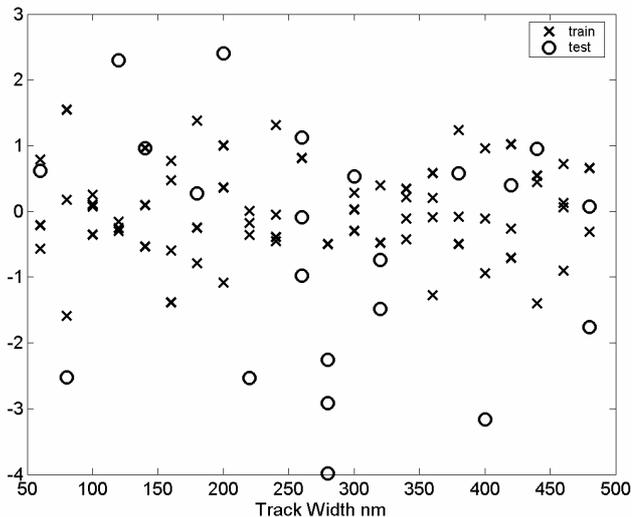


Figure 3. A typical result showing the spread of measurements, unit of vertical axis in nm.

3. SUMMARY OF PREVIOUS RESULTS

The system described in section 2 has been used to study different samples, some of the results have been published elsewhere [12] and others are to be published. In this section we will use a couple of these results to demonstrate the abilities of the technique.

Sample 1: single tracks measurement The sample was a silicon substrate with isolated tracks etched on it. The track widths ranged from 60 nm to 480 nm in steps of 20 nm. The optical system used a 0.3 NA/x10 Zeiss Epiplan objective and a HeNe laser. This corresponded to a psf of 2.57 μm , and a lateral resolution of 1.29 μm according to the Rayleigh criterion. Each track was scanned several times and the profiles were processed with a trained ANN. Figure 3 shows the spread of the measurements. A number of observations can be made:

1. the standard deviation of the spread is 1.8 nm, which is about 1/1400 of the diameter of the psf of the system;
2. the smallest track in the set is 60 nm, which is 1/43 of the diameter of the psf;
3. 60nm does not represent the limit of the system set up. It is merely the smallest sample available;
4. by using a shorter wavelength light source and a higher NA objective lens, a track width of 10 nm should be within the measurement range of the system;
5. a number of ANNs were trained under different conditions, and similar results were obtained;
6. tracks that were outside the training ranges of the networks would result in **large** error;
7. the measured tracks did not need to coincide with any of the reference samples.

Sample 2: double tracks measurement The sample consisted of pairs of chromium tracks deposited on a glass substrate. The widths of the tracks ranged from 1 to 3 micrometres and separations 1 to 4.8 micrometres. The pairs of tracks were well separated from each other. The NA of the objective used was reduced to 0.11. Two networks were tested. The first was a single ANN that consisted of two outputs: width and separation. The second network consisted of two ANNs, one measured the width and the other the separation. The results are shown in Table 1. It is clear that the two ANN network performed better than the single one, the percentage spread ($\sim 0.14\%$) is two times larger than the case of single track measurement. The last point is of particular interest and will require further investigation.

| | One ANN two outputs | | Two ANNs each one output | |
|-----|---------------------|--------------|--------------------------|--------------|
| | Width nm (%) | Sep nm (%) | Width nm (%) | Sep nm (%) |
| STD | 25.4 (0.36%) | 14.5 (0.21%) | 9.52 (0.14%) | 9.29 (0.13%) |

Table 1: Double tracks, STD of spreads from one ANN with two outputs, and two ANNs each one output. Figures in brackets are percentages with respect to the psf

4. EFFECTIVE BANDWIDTH AND NOISE AMPLIFICATION

The results shown in section 3 demonstrated the capability of the technique. We will now concentrate on the objectives of

this paper and address the theoretical background of the method. From the experimental results shown above, it is clear that the system needs to be extremely repeatable and possess high signal to noise ratio in order to provide the measurement accuracy. These requirements, however, are normal for any good measurement instrument. The optical system we used is constructed on an optical breadboard mounted on an optical table. The system is located in a general purpose laboratory without any special isolation facility. Even better measurement results could be expected if the environmental conditions are improved. These requirements, although necessary, are not sufficient to guarantee good results. In the next three sections, we will examine factors that will critically affect the system.

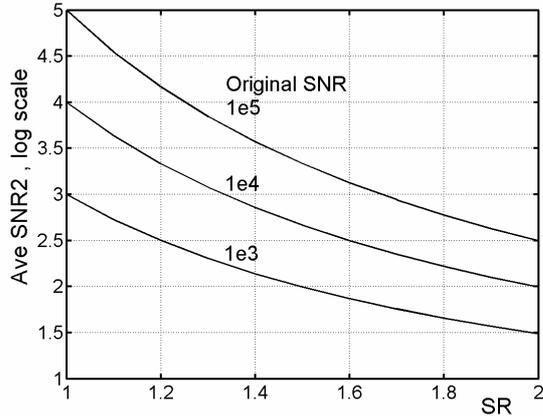


Figure 4. Reduction in system SNR for three different starting SNR. System NA = 0.3 and $\lambda = 633\text{nm}$.

As mentioned in section 1, information theory can be used to describe the output of an optical system, and the information content N of a profile obtained from a time independent 1D system is given by

$$N = (1 + 2LB) \log_2(1 + s/n) \quad (1)$$

where L is the length of the profile, B is the bandwidth of the system, and s/n is the average signal to noise ratio of the profile. Once a profile is taken, the value of N is fixed and cannot be increased through processing of the profile. The effective bandwidth of the signal can be increased but only at the expense of other parameters. Figure 4 is a plot showing the reduction of the average signal to noise ratio as a function of SR , where SR is the bandwidth extension factor given by

$$SR = \frac{\text{bandwidth after extension}}{\text{bandwidth before extension}} \quad (2)$$

Figure 4 is obtained with a system NA of 0.3, and a wavelength of 633nm. The three initial system SNR have been shown and the reductions in the SNR values are severe. The effects of these reductions on the measurement precision will be examined.

We will use the measurement of track widths as examples to investigate the relationships between the effective bandwidth, noise and the measurement precision. We will calculate the change in the sample profile as a function of the track width, and compare it to the noise level of the system, before and after the extension of the system bandwidth. Briefly, the process involves is [14]

- A Type I [15] scanning optical microscope is modelled to give the intensity profiles of tracks of widths w_i . The

optical model will not be discussed further as it does not affect the investigation;

- A particular number of detected photons, is assigned to the profile. This will define the SNR of the profile if photon shot noise is taken as the only significant noise source;
- The profile is integrated to give the total number of photons, P_i ;
- The process is repeated with different values of w_i .

Figure 5a is a plot of P_i against w_i . The optical system modelled has the same parameters as the one used in figure 4, and the averaged number of detected photons is 1×10^6 per scan point. Next we differentiate the curve $P(w)$ and calculate the parameter R :

$$R = \frac{\partial P}{\partial w} \delta w \quad (3)$$

Figure 5b shows R as a function of w , with $\delta w = 2$ nm and an averaged number of detected photons of 5×10^7 per scan point. Also shown is the average noise level of the system. The curve R intersects the noise level at $w = 34$ nm, meaning that it is capable of a measurement precision of 2 nm at the track width of 34 nm.

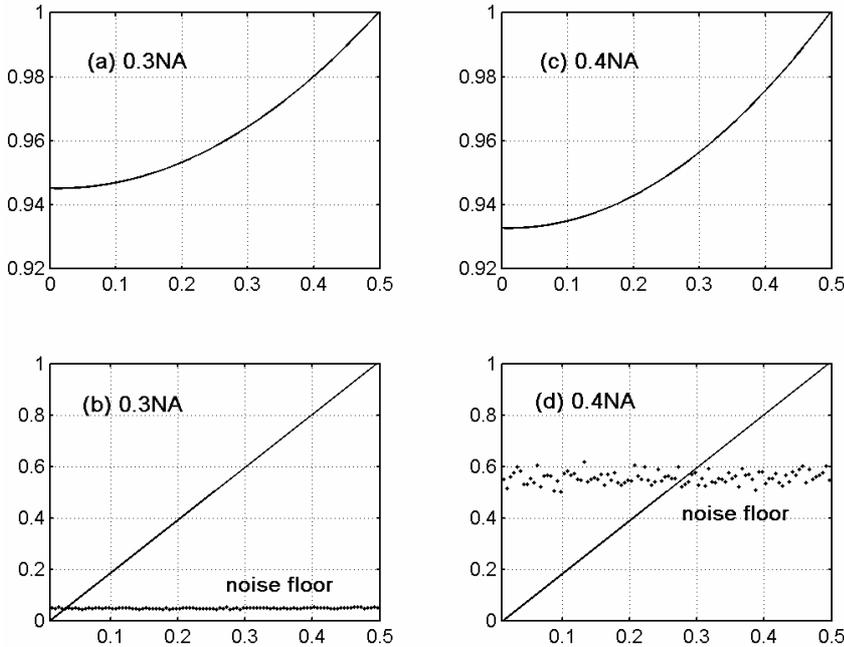


Figure 5. Effects of noise on measurement precision, a & b: system NA = 0.3; c & d: effective NA increased to 0.4.

Horizontal axes: track width in micron
Vertical axes: relative intensity

* b & d have been normalised for ease of comparison.

The simulation is repeated with an effective system NA of 0.4, or $SR = 1.33$. The reduced SNR, according to eq. (1), corresponds to a detected number of photons of 6×10^5 per scan point. Figures 5c & 5d are the resulting curves. The R curve now intersects the noise floor at the track width of 280 nm, which is much worse than the first case. This result may appear to be counter intuitive, but is merely a reflection that, although increasing the effective bandwidth will improve the system sensitivity, is actually increases the effects of noise much more. This point is illustrated further at figure 6, where the intersection level (IS) is plotted as a function of the average number of detected photons per scan point. Two system bandwidths, 0.3 and 0.4 NA have been simulated. Considering point A on the 0.3NA curve, it has an IS value of 70 nm

with an average number of detected photons of 2×10^7 . If we change the NA of the optical system to 0.4 but keeping the SNR unchanged, the operating point will be shifted to B, which gives an IS value of 62 nm. If, however, the bandwidth is increased by means of signal processing of the profile generated with the 0.3NA system, eq. (1) will apply. The averaged NOP of the resulting system would decrease to 4.3×10^4 , which is indicated by line C in figure 6. As can be seen, line C does not intersect curve B. This means that, with this level of SNR, the system is not capable of measurement track width with a repeatability better than 2 nm (at least for tracks smaller than $0.5 \mu\text{m}$)

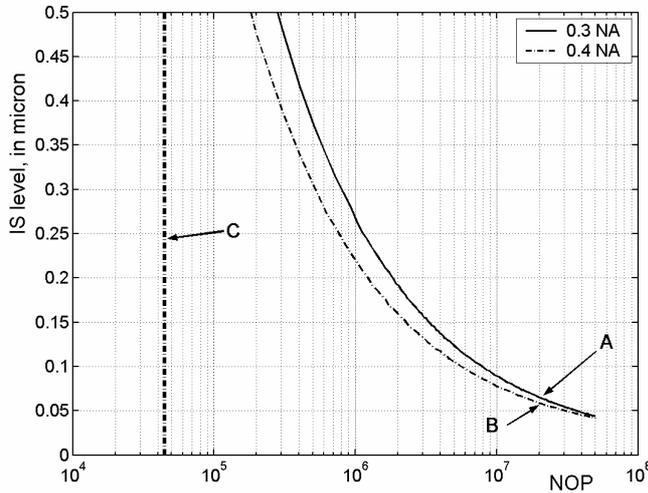


Figure 6. Relationships between extending the system effective NA, the signal to noise ratio of the system, and the measurement precision

To further illustrate this point, figure 7a shows the profile of a track obtained using a 0.3 NA system without addition of any noise. Figure 7b is the profile if the effective bandwidth of the system NA is extended to 0.6. Again, no noise has been imposed. If we now impose noise onto the profiles, using the relationship stated in eq. (1), figures 7c & d are resulted. The average number of photons per scan point is 5×10^7 in 7c, this yields an average NOP of 7×10^3 per scan point. It is not surprising that the resulting level of noise would severely affect the measurement precision.

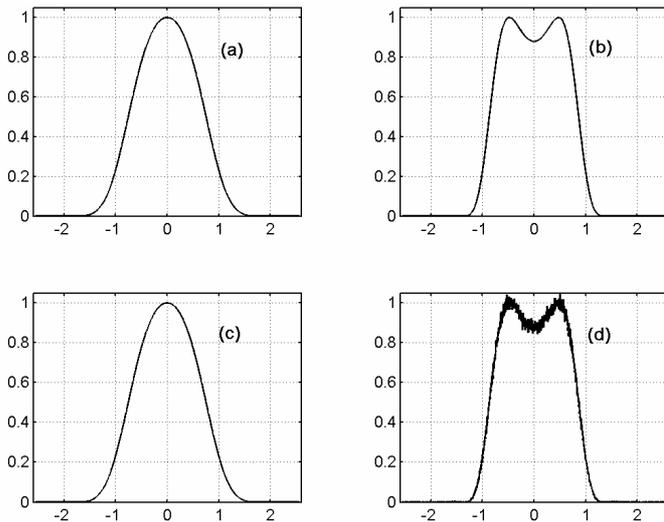


Figure 7. Effects of extending the effective bandwidth on the noisy profiles.

5. ANN AND OPTIMUM DATA FILTERING

The results presented above indicate that extending the effective bandwidth does not actually improve the measurement precision. Improvement can be had, however, by using data selected specifically from the profiles. This is basically the essence behind our technique. With reference to the signal conditioner (SC) in figure 1, it has two functions: 1) to calculate the spectra of the differentials of the optical profiles; and 2) to select a number of points from the differential spectra for the ANN inputs. The first point has been discussed elsewhere [16] and will not be repeated here, suffice it to say that the ANNs failed to function when other data format was used. The second point is directly relevant to the discussion here. We have found that both the number of spectral points and their locations can critically affect the measurement. The experimental results shown in section 3 were obtained using eight points, spaced equally in the frequency domain. This format produces good results although it is still not optimum. It was found subsequently that by reducing the number of inputs to five and avoiding the low frequency range of the spectrum, slight improvement can be achieved [16]. Conversely it was found that some regions of the spectrum could have negative influence on the measurement, and their inclusion could make the results worse. Moreover, the format of this 'ideal' pattern is dependent on the sample under examination. An isolated track and two closely spaced tracks, for example, will require different input formats if best results are to be obtained. All these are important considerations when designing the signal conditioner and the ANN. Further research is required and the findings will be presented in a future publication.

6. SIGNAL CLASSIFICATION AND MULTIPLE ANNS

Using the system described, we have measured different samples with high degree of precision. We have also shown, through computer simulations, that other structures, such as tracks with asymmetric sidewall angles, could be measured successfully. Several important factors, such as good system stability and high signal to noise ratio, and appropriate signal conditioning, have been discussed above. Another important factor is that complicated measurement tasks should not be placed upon the ANNs, or their performance would be degraded. This point is clearly demonstrated in section 3, where two ANNs performed much better than the single ANN.

If the system is required to carry out different types of measurements, many ANNs should be used with each trained for one specific measurement. This should not present much difficulty, as the training and the storage of the ANNs are relatively straightforward. Matching the sample with the appropriate ANN, however, can be problematic if *a priori* knowledge of the sample does not exist. To make the system even more powerful, it will need to be able to recognise the sample and determine the appropriate ANNs to use. This calls for classification of the samples and will be discussed briefly here.

Many techniques exist for classifying object and self-organising map is a good example [13]. We have also trained a feed forward ANN (figure 2) for the task. To test the network, the single and double tracks profiles described in section 3 were mixed together. When the ANN was applied to the profiles, the single and the double tracks were separated without any failure.

Another group of techniques that can potentially be used for this purpose is the multivariate analysis algorithms, among

them are the principal component analysis, independent component analysis, factor analysis and projection pursuit [17]. Initial study of this group of techniques has shown great potential for our application, and we will use the principal component analysis (PCA) to describe briefly the principle.

PCA is an orthogonal decomposition technique using eigenfunction analysis. The aim is to project data onto a set of mutually perpendicular axes. The resulting eigenvalues and eigenvectors indicate the degree of significance of different components. Imagine a set of two dimensional data given by

$$X = [x_1, x_2, x_3, \dots, x_n] \quad \text{and} \quad Y = [y_1, y_2, y_3, \dots, y_n] \quad (4)$$

The first step of PCA is to calculate the covariance matrix, $C_v\{X,Y\}$, of the data. Next, eigenvalue decomposition is applied to C_v , yielding two eigenvectors with two eigenvalues. If one eigenvalue is much greater than the other one, it means that most of the energy contained in the data is concentrated along the direction given by the corresponding eigenvector. For our analysis, we need to have a set of reference profiles R , generated using the optical system. The reference samples should contain the different structures of interest. If S is the unknown profile to be measured, it can be treated as the X -vector in (3). Selecting one reference profile from R , PCA can be applied to R - S . The process is then repeated with the other reference profiles. Or the entire set of references can be used at once, as PCA is used for multi-dimensional analysis. This will yield an array of eigenvalues and eigenvectors. By examining this array, much information regarding the sample can be gained.

Classification is not a one step process. Depending on the complexity of the sample, several iterations with different PCA and ANN may be needed. As with other forms of search and matching algorithms, the technique will work best if the sample range is limited. The success or otherwise of the classification depends on the availability of the reference profiles, which can be added to the system when they become available. In other words, the application of the system can grow and keep pace with the demand of the user.

Another potential advantage of the PCA is that it can simplify the profiles before they are input to the ANNs. In addition to the eigenvalue and eigenvector, we can extract the eigenfunction of the principal component as well. The eigenfunction is a least mean square best match of the reference profile and is particularly useful for complicated samples, as most of the irrelevant information is removed from the eigenfunction and will ensure good performance of the ANNs.

7. CONCLUSIONS

In this paper, we have presented a technique for precise measurement of structures that are in the nanometre scale. The system can be divided into three sections: optical instrument, signal classification and conditioning, and signal measurement. The optical section and the signal measurement section are relatively straightforward, and their workings have been demonstrated with experiments. Track width measurement down to 60 nm with uncertainty less than a couple of nanometres has been achieved. Our existing system consists of the optical section, the signal conditioning, and the

measurement section. If the application of the system is to be broadened, the classification section will become crucial. Much future work is required for this section in order to determine which type of multivariate analysis is suitable for the sample type concerned. Our effort has concentrated on measuring samples of different structures. Another important goal is to measure multiple parameters of a sample, as it will open up the possibility of optical imaging with resolution beyond the diffraction limit. Findings of this work will be presented in future publications.

ACKNOWLEDGEMENTS

We would like to thank the Engineering and Physical Science Research Council (UK) and the National Physical Laboratory (UK) for supporting this research. Part of this work was carried out under the DTI Programme for Length Measurement 2002-2005.

REFERENCES

1. G. Toraldo Di Francia, "Resolving power and information," *J. Opt. Soc. Am.* **45**, 497-501 (1955).
2. H. Wolter, in *Progress in Optics*, E. Wolf, ed. (North-Holland, Amsterdam, 1961), Vol. 1, Chap. 5.
3. J. L. Harris, "Diffraction and resolving power," *J. Opt. Soc. Am.* **54**, 931-936 (1964).
4. C. W. Barnes, "Object restoration in a diffraction limited imaging system," *J. Opt. Soc. Am.* **56**, 575-578 (1966).
5. R. W. Gerchberg, "Super-resolution through error energy reduction," *Opt. Acta* **21**, 709-720 (1974).
6. D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty – I," *Bell Syst. Tech. J.* **40**, 43-63 (1961).
7. H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty – II," *Bell Syst. Tech. J.* **40**, 65-84 (1961).
8. C. K. Rushforth, and R. W. Harris, "Restoration, resolution, and noise," *J. Opt. Soc. Am.* **58**, 539-545 (1968).
9. W. Lukosz, "Optical systems with resolving powers exceeding the classical limit," *J. Opt. Soc. Am.* **56**, 1463-1472 (1966).
10. I. J. Cox, and C. J. R. Sheppard, "Information capacity and resolution in an optical system," *J. Opt. Soc. Am. A* **3**, 1152-1158 (1986).
11. N. B. E. Sawyer, C. W. See, M. Clark, M. G. Somekh, J. Y. L. Goh, "Ultrastable absolute-phase common-path optical profiler based on computer-generated holography," *Appl. Opt.* **37**, 6716-6720 (1998).
12. R. J. Smith, C. W. See, M. G. Somekh, A. Yacoot, and E. Choi, "Optical track width measurements below 100 nm using artificial neural networks," *Mea. Sci. Tech.* **16**, 2397-2404 (2005).
13. S. Haykin, *Neural Networks – A Comprehensive Foundation* (2nd ed., Prentice Hall, New Jersey, 1999).
14. J. Nunn, W. Mirande, H. Jacobsen and N. Talene 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, 53–68 (1997).
15. T. Wilson, and C. J. R. Sheppard, *Theory and practice of scanning optical microscopy*, (Academic Press, London, 1984).

16. R. J. Smith, *Ultra-fine optical line width measurement using artificial neural networks*, PhD thesis, University of Nottingham, August 2006.
17. Lindsay Smith, “A tutorial on principal components analysis”,
http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf