Use of artificial neural networks on optical track width measurements

Richard J. Smith,[1,*] Chung W. See,[1] Mike G. Somekh,[1] and Andrew Yacoot[2]

[1]School of Electrical and Electronic Engineering, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom

[2]Industry and Innovation Division, National Physical Laboratory, Hampton Road, Teddington, Middlesex TW11 0LW, United Kingdom

*Corresponding author: eexrjs@nottingham.ac.uk

Received 20 March 2007; accepted 17 April 2007;
posted 24 April 2007 (Doc. ID 81276); published 26 June 2007

We have demonstrated recently that, by using an ultrastable optical interferometer together with artificial neural networks (ANNs), track widths down to 60 nm can be measured with a 0.3 NA objective lens. We investigate the effective conditions for training ANNs. Experimental results will be used to show the characteristics of the training samples and the data format of the ANN inputs required to produce suitably trained ANNs. Results obtained with networks measuring double tracks, and classifying different structures, will be presented to illustrate the capability of the technique. We include a discussion on expansion of the application areas of the system, allowing it to be used as a general purpose instrument. © 2007 Optical Society of America

OCIS codes: 120.3180, 120.3940, 120.4640, 200.4260.

1. Introduction

In a recent publication [1] we introduced what we believe is a novel technique for optical linewidth measurement. The technique makes use of an ultrastable scanning optical interferometer to generate profiles of the samples. Artificial neural networks (ANNs) are then applied to the profiles to extract the desired parameters: track width and track separation. Using an optical setup with a point spread function of 2.6 μm in diameter (NA = 0.3, λ = 0.6328 μm), we successfully measured linewidths down to 60 nm, with a repeatability better than 2 nm. We will present new results that show the characteristics and capability of the system. In particular, we will demonstrate the flexibility of the technique and show that it can provide accurate measurements under various experimental conditions. We will also consider aspects that are critical for the future development of the technique.

Optical techniques have been the mainstay of critical dimension (CD) measurements for many years [2]. Optical systems are noncontact and nondestructive. They are reliable, easy to use, and relatively inexpensive. The use of monochromatic illumination enables measurements that can be related traceably to the international recognized standard of length, the meter. All these factors have made them the instruments of choice for CD measurements. The situation, however, has changed dramatically in recent years. Due to advances in semiconductor technology, feature sizes have decreased rapidly and structure dimensions below 100 nm are becoming common. These small features are beyond the capabilities of conventional optical systems, and techniques such as atomic force microscopy (AFM) and scanning electron microscopy (SEM) are now regularly employed for these measurements. Although their lateral resolutions are not matched by optical systems, they do not possess many of the attributes associated with the optical systems: noncontact, direct traceability, and nondestructive. Therefore, there has been much research on extending the measurement capabilities of optical systems.

Many approaches have been proposed to tackle this problem, with the aim of increasing the effective bandwidth of the system [3–6]. The amount of improvement achieved in practice, however, has been negligible. Al-

0003-6935/07/224857-10$15.00/0

though the suggested techniques have a solid mathematical basis, there are also obstacles that are difficult to overcome.

Consider an optical imaging system, its output $v(x)$ related to the input $u(x)$ and the system impulse response $h(x)$ through the convolution integral

$$\begin{aligned} v(x) &= \int_{-\infty}^{\infty} \mathrm{rect}\left(\frac{\chi}{T}\right) u(\chi) h(x-\chi)\mathrm{d}\chi + n(x) \\ &= \int_{-T/2}^{T/2} u(\chi) h(x-\chi)\mathrm{d}\chi + n(x) \\ &\textit{and } \mathrm{rect}(x) = 1 \quad \textit{for } |x| \le \frac{1}{2} \\ &= 0 \quad \text{otherwise}, \end{aligned} \tag{1}$$

where T is the width of the object, and $n(x)$ the noise associated with the image. A one-dimensional (1D) amplitude system is shown for simplicity. To extract object information originally outside the bandwidth of the microscope, two steps need to be taken: (1) deconvolve the output $v(x)$ from the impulse response, and (2) extend the resulting spectrum beyond the original system bandwidth. The main concern regarding the deconvolution is the amplification of noise, which is inevitable if an inversion filter is used. A Wiener filter, although preserving the signal-to-noise ratio (SNR) of the original image, will result in a modified transfer function, thus leading to error in the extended spectrum [7]. Regarding spectrum extension, an important consideration is what we believe is the uniqueness of the extended spectrum. Toraldo di Francia [8] showed that it was possible for two different objects to produce identical images (and hence identical spectra). This ambiguity could be removed, Toraldo argued, if prior knowledge of the object was available [9]. Wolter [10] and Harris [3] showed that this knowledge could be the object size being finite. This is because the Fourier spectrum of a spatially bounded object is an analytic function, and by applying the principle of analytic continuation [11], we believe that a unique extension of the spectrum can be obtained.

Many researchers have investigated the effects of noise on the extension of the effective bandwidth. One model makes use of the information theory [12–14] and concludes that the information content N associated with any particular object profile is fixed, with its value proportional to the number of data points, the spatial (and temporal if appropriate) bandwidth, and the SNR of the system. The profile may be processed to trade one parameter with another, but the value of N cannot be increased. Although the effective bandwidth of a profile may be extended beyond the classical limit, it can only be so at the expense of other parameters. In a series of publications, Slepian and co-workers [15–17] demonstrated that Eq. (1) can be represented using the prolate spheroidal wave functions. Based on these functions, Rushforth and Harris [18] showed that the error of the extension process increased nonlinearly with the number of eigenfunctions included in the reconstruction, and regardless of the characteristic of the optical system, large noise amplification would result even for a small extension of the system bandwidth. Many different methods have been proposed for extending the bandwidth, but the actual extension achieved has been extremely limited. One successful implementation of spectrum extension is in Fourier transform infrared spectroscopy. It was reported that, by using an autoregressive model and a singular value decomposition, an eightfold increase in the signal length was achieved [19]. This, perhaps, is not surprising, as the signal in question consisted of a small number of sine waves only.

We tackle this problem from a different angle. Instead of attempting to extend the effective bandwidth of the system, we aim to measure accurately the dimensions of the sample. Based on these measurements, we can build up a clear picture of the sample. The advantage of this approach is that we do not need to perform deconvolution of the signal, thus avoiding amplification of the system noise. In addition, the measurements of individual parameters of the object do not require full extension of the bandwidth, thus avoiding the severe degradation in the SNR of the system.

In Section 2 we will briefly recap the technique we reported previously. This will be followed by the presentation of results, showing both the measurement capability and the characteristics of the system. In Section 4 we will discuss future work necessary to allow the system to be used as a routine measurement instrument, and also to increase its application areas. Our conclusions are presented along with an Appendix.

2. Optical System and ANNs

The technique can be divided into two parts: an optical section that is capable of producing extremely repeatable profiles of very high SNR, and a signal processing section that is constructed specifically for our application. We will discuss the signal processing algorithm first, as its requirements will define the specifications of the optical system.

The signal processing technique we use is the ANN. It is a feed forward single layer perceptron [20]. Networks of this type have been shown to be especially suitable for extracting small changes in the input signal [21]. The measurement concept is simple: the optical system will provide a profile of the sample under test. The profile, after suitable signal conditioning, is fed to an ANN that has been trained for the particular application. The output will be the object parameter desired.

The training process involves repeatedly exposing profiles of known objects to the ANN. At each cycle, the weights of the nodes are adjusted to reduce the errors between the network outputs and the target values of the profiles. Figure 1 shows the training of the ANN in block diagram form. Details relating to

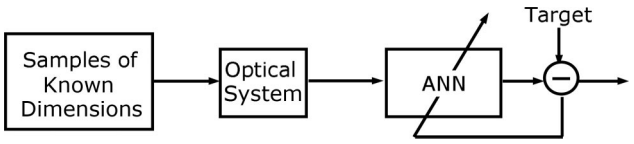

Fig. 1. Training of ANN using samples of known dimensions and the back propagation algorithm.

the training of the ANN for our applications can be found in Ref. 1 and are summarized in Appendix A.

The important properties that the optical system should possess are high stability, high SNR, and the ability to produce both amplitude and phase profiles. The first two are obvious and are properties demanded for good optical instruments. The third property would allow the system to examine different types of samples, such as chrome on glass structures or etched features on homogeneous substrates. The system we use is one that we developed previously [22], which is a scanning interferometer. It makes use of a computer generated hologram as the beam splitter, which allows the two interfering beams to travel along a similar path. The effects of microphonics and thermal gradients of the surrounding are therefore greatly reduced. Due to the special detection method used in the system, the interferometer has a very high SNR. The number of detected photons per scan point can reach 2×10^9 and is ideal for the present application. In the next section, we will present results showing the measurement capability of the system, and also demonstrate the characteristics of the technique.

3. Results

The experimental results presented here were obtained using the scanning interferometer described above. The objective was a $10\times/0.3$ NA Zeiss Epiplan and the laser was a helium–neon (633 nm). The diameter of the point spread function (PSF) at the sample surface was 2.6 μm. The sample stage was a Physik Instrumente P-517.3 piezoflexure nanoscanner with capacitor position sensors and had a resolution of better than 1 nm. The measurement procedure has been described in a previous publication [1] and will not be repeated here. A number of samples were used in the experiments, and their specifications are given in Table 1. The results can be divided into two groups. The first will show the measurement capability of the technique, and the second will illustrate important characteristics of the system.

A. Track Width Measurements

Two types of feature were tested: single isolated tracks and double tracks of various widths and separations. Single track results similar to those presented here have been published before [1], although they were obtained on different samples. The new measurements presented here show better repeatability compared to previously reported results with an improvement approaching a factor of 2. Apart from those contained in Subsection 3.A.1, results presented in this paper have not been published before.

1. Single Tracks

Two samples were used. The first (Community Bureau of Refere sample) comprised a series of chrome lines on glass and was a standard produced as part of an international collaboration between national measurement institutes [2]. Lines with widths in the range from 0.27 to 2.1 μm were scanned. The second sample was a silicon sample with a series of tracks with widths in the range from 60 to 480 nm etched in its surface. As well as covering a different range of track widths, the samples also allowed us to test the system performance on amplitude and phase data. Details of the samples used are given in Table 1.

The two samples were scanned with the optical interferometer. According to the procedure described in Appendix A, each track was scanned four times to make up the training and testing samples. Two ANNs were trained, one for each sample. Figure 2 shows the differences between the ANN outputs and the targeted track widths. The crosses represent data from the training set and the circles the testing set.

The measurements are summarized in Table 2. The repeatability is the standard deviation calculated for the entire set. The smallest track, being 60 nm, is less than $1/40^{\text{th}}$ of the focal spot diameter. The repeatability is even more impressive, being smaller than $1/1000^{\text{th}}$ of the PSF. It should be noted that the 60 nm track does not represent the limit of the system and is merely the narrowest track available to us. Indeed, the samples were measured using a $5\times/0.13$ NA objective, and results of comparable repeatability were obtained.

These two samples were measured repeatedly under different conditions. Similar results were obtained each time. Based on these measurements, we predict that, by using an objective of higher NA and a laser of shorter radiation wavelength, measurement capability for linewidths in the 10 nm region should be achievable.

Table 1. Samples Used in the Experiments

Sample	Type	Track Widths	Comments
Si	Single tracks, edged Si substrate	60–480 nm	Track separation: 60 μm, width not calibrated
BCR	Single tracks, Cr on glass	0.27–2.1 μm	Track separation: 20 μm, width calibrated by NPL
Chrome	Single and double tracks, Cr on glass	1–3 μm	Separation of double tracks: 1–3 μm

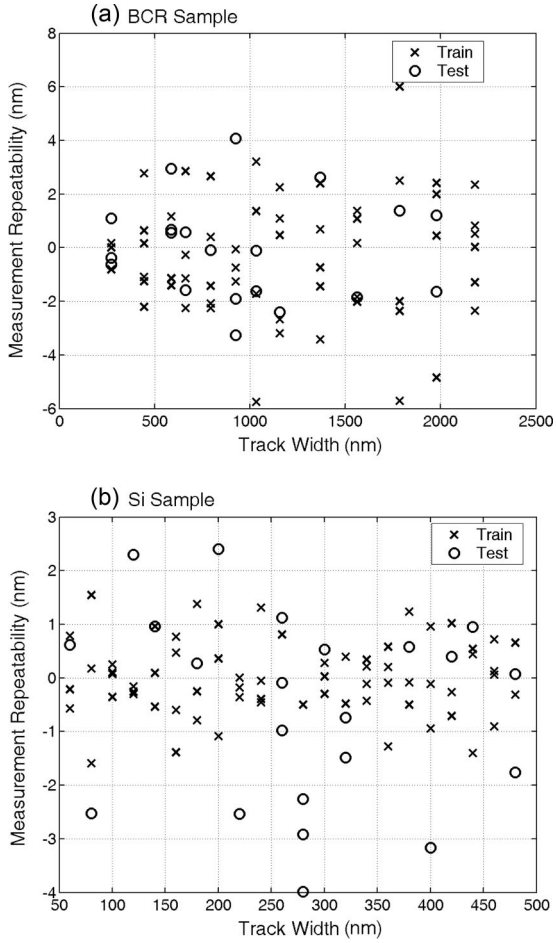

Fig. 2. Measurement repeatability: (a) BCR sample and (b) Si sample.

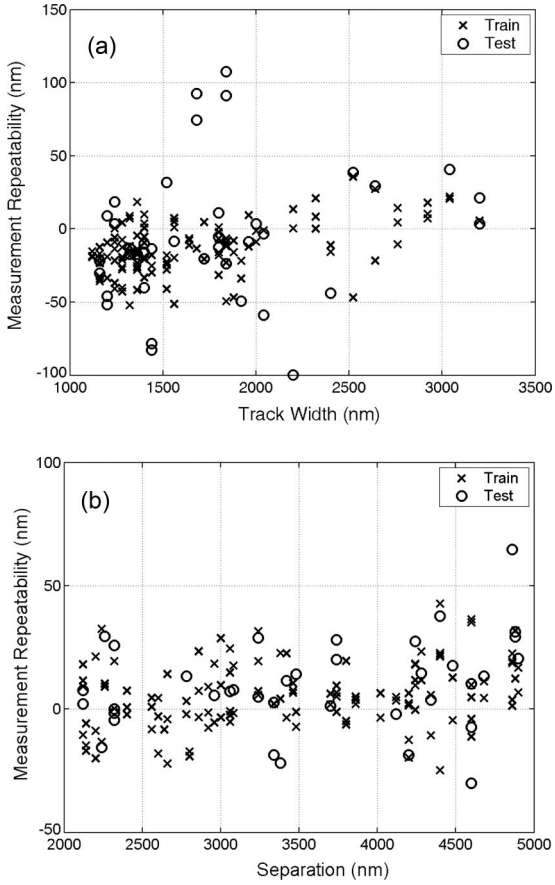

Fig. 3. Measurement repeatability of double tracks. One ANN with two outputs: (a) track widths and (b) track separations.

2. *Double Tracks*

The chromium sample that contained double tracks was measured. The widths of the individual tracks ranged from 1 to 3 μm with separations up to 4.8 μm. Each double track was measured three times and in total there were 162 double tracks to use for training. Because of the relatively large dimensions of the features, the NA of the objective lens was stopped down to 0.11. Two networks were trained: the first was a single network containing two outputs, one for the widths and the other for separations; the second contained two ANNs, one trained to measure width and the other separation. The objective of this exercise was to investigate the ability of the ANN in measuring multiple parameters.

Figure 3 shows the results obtained with the single network with two outputs, and the results are summarized in Table 3. As can be seen from the figure, there were a number of poorly performing tracks in the testing set. This is probably because although there were many samples, they covered a large range. The samples that were poorly measured corresponded to regions that were inadequately covered.

The same set of samples was measured again, but this time with two ANNs, one trained specifically for track widths and the second for separations. The results are included in Table 3. Compared to the previous case, much better repeatability is shown. These results would be improved further if the number of

Table 2. Measurement Repeatability for the BCR and Si Samples

	Measurement Repeatability in nm, Standard Deviation			
Sample	Training	Testing	Spread wrt psf	Smallest Track wrt psf
BCR	2.2	1.9	1:1370	1:10
Si	0.7	1.8	1:1440	1:43

Table 3. Measurement Repeatability for Double Tracks, Showing Percentages of Point Spread Function

	One Net Two Outputs		Two Nets Each One Output	
	Width nm (%)	Sep nm (%)	Width (%)	Sep (%)
Std train	17.3 (0.25)	13.4 (0.19)	5.20 (0.07)	3.42 (0.05)
Std test	25.4 (0.36)	14.5 (0.21)	9.52 (0.14)	9.29 (0.13)

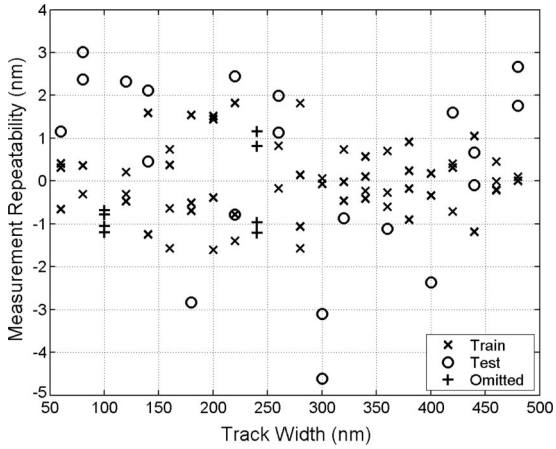

Fig. 4. Measurement repeatability with two tracks (100 and 240 nm) omitted from the training set.

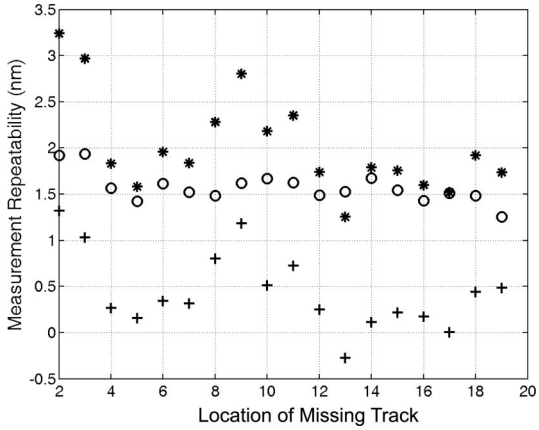

Fig. 5. Effects of the location of the omitted track on the measurement repeatability. "*": error associated with the omitted track; "○": average error of the entire network; and "+": the difference between the two.

examples of width and separation included in the training set was increased.

B. System Characteristics

The single track results shown above were obtained with networks trained under ideal conditions, in the sense that the training samples were of known dimensions and were uniformly spaced over the measurement range. Next we examined the behavior of the networks when this is no longer the case.

1. Missing Tracks Left Out at Random

To simulate the situation where the training samples did not uniformly cover the measurement range, we used the Si sample results from Subsection 3.A.1 and randomly left out two tracks from the training set. After the training, the network was tested with the full data set, including the tracks omitted originally. From the result shown in Fig. 4 it is clear that the errors introduced by omitting the tracks (100 and 240 nm) are no greater than for the rest. This shows that the network produces a general model of the input–output relationship, and that the ANN should be able to measure samples that are not originally in the training set.

Further tests were carried out to investigate the effects of missing tracks. These included the omissions of one, two, and more tracks from the training. All tests were performed using the results obtained from the Si sample as described in Subsection 3.A.1. The findings are summarized as follows:

Omitting one track: The point of this exercise was to examine the effects of the locations of the missing tracks on the ANNs. The procedures of the simulations were:

- Poisson noise was added to each of the 21 profiles, so that the noise power was double that of the original levels;
- A network was trained with track 2 (80 nm wide) omitted;
- The network was then applied to all 21 tracks, and the repeatability for each track was recorded;
- The above was repeated 200 times, each with different sets of Poisson noise, resulting in a distribution of the repeatability;
- The entire process was repeated for tracks 3–20.

The addition of the random noise was to ensure that the outcomes were general. The results are presented in Table 4, with the top row showing the location of the missing track; the second row showing the repeatability of the measurements over the 200 runs for the particular missing track; the measuring third row is the repeatability of all 21 tracks for each run, then averaging them over the 200 runs; and the fourth row containing the differences between the second and the third rows.

From the results shown in Fig. 5 one can observe that the variation over the range is relatively small, indicating that the effects of the locations of the missing tracks are not significant.

Omitting two or more tracks: This was similar to the last case, but for two or more missing tracks. In this case, 500 runs were used instead of 200, and for each run the locations of the omitted tracks were selected randomly. For each number of omitted tracks, the errors were averaged. The results obtained are shown in Fig. 6 and Table 5. Not surpris-

Table 4. Effects of the Locations of Missing Tracks on the ANNs. See Section 3.B.1

N_{mt}	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
SD_t	0.216	0.198	0.122	0.105	0.131	0.123	0.152	0.187	0.145	0.157	0.116	0.084	0.119	0.117	0.107	0.101	0.128	0.116	0.114
ME_c	0.14	0.14	0.115	0.104	0.118	0.111	0.107	0.12	0.123	0.12	0.11	0.106	0.121	0.105	0.105	0.108	0.106	0.087	0.102
Dif	0.076	0.058	0.007	0.001	0.013	0.012	0.045	0.067	0.022	0.037	0.006	−0.022	−0.002	0.012	0.002	−0.007	0.022	0.029	0.012

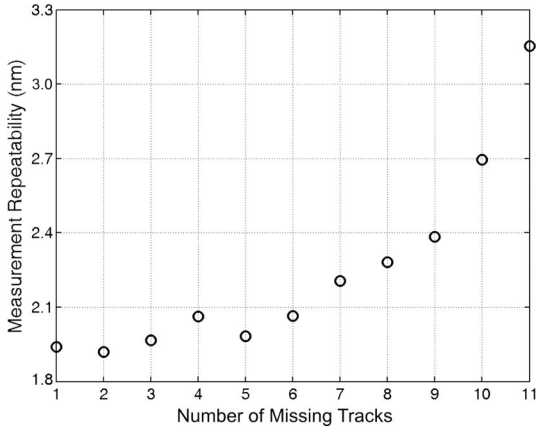

Fig. 6. Effects of number of omitted tracks on measurement repeatability.

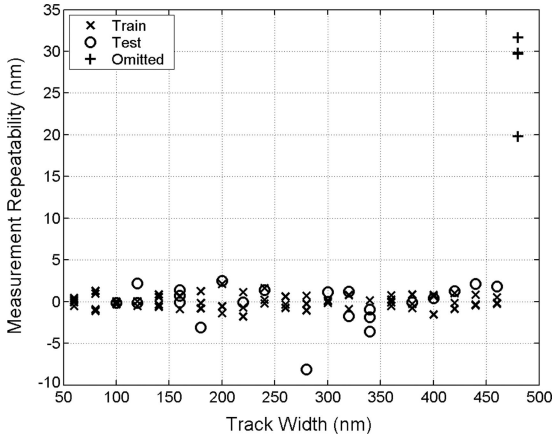

Fig. 7. Measurement of sample outside of the training range.

ingly, the errors increase as does the number of omitted tracks. Note that the average values have been taken over 480 runs instead of 500. This is because not all the networks could be trained successfully. For cases of a small number of omitted tracks, only one or two networks failed to train. For the case of 11 omitted tracks, 15 networks failed. They were removed before the averaging was taken. Another point observed from the test was that the failed networks usually occurred when the omitted tracks were grouped together rather than being spread evenly over the range.

C. Out of Range

The networks that have been trained will only be valid for tracks that fall within the ranges. This is demonstrated in Fig. 7, where the network was trained on the range of 60–460 nm, and then the trained network was applied to the 480 nm tracks. The error, both the mean and the standard deviation, for the 480 nm tracks was considerably higher. Similar results were obtained when the 60 nm tracks were omitted from the training. These results, together with the previous one on omitting tracks inside the training range, demonstrate that these networks are very good at interpolation across the training range but are unsuitable for extrapolating outside the training range.

D. Input Point Locations

The format of the input signal to the ANN is critical to the performance of the network. Both the normalized object profiles and the spectra of the profiles were used as the inputs to the ANNs. Although networks were trained successfully, the measurement spreads were greater than 50 nm. The inputs used for the ANNs shown above are the spectra of the *differential* profiles of the objects, and the repeatability is now typically around a couple of nanometers. It can be argued that, for measuring widths and separations, low spatial frequency components have little to contribute. By differentiating the profiles, the effects of the mid- to high-frequency components are enhanced.

Figure 8(a) shows a typical input to the network, which has 36 sample points within the passband of

Table 5. Measurement Repeatability (MER) (over 480 Runs) as Functions of the Number of Tracks (NTO) Omitted

NTO	1	2	3	4	5	6	7	8	9	10	11
MER (nm)	1.9	1.9	1.9	2.0	2.0	2.0	2.2	2.3	2.4	2.7	3.1

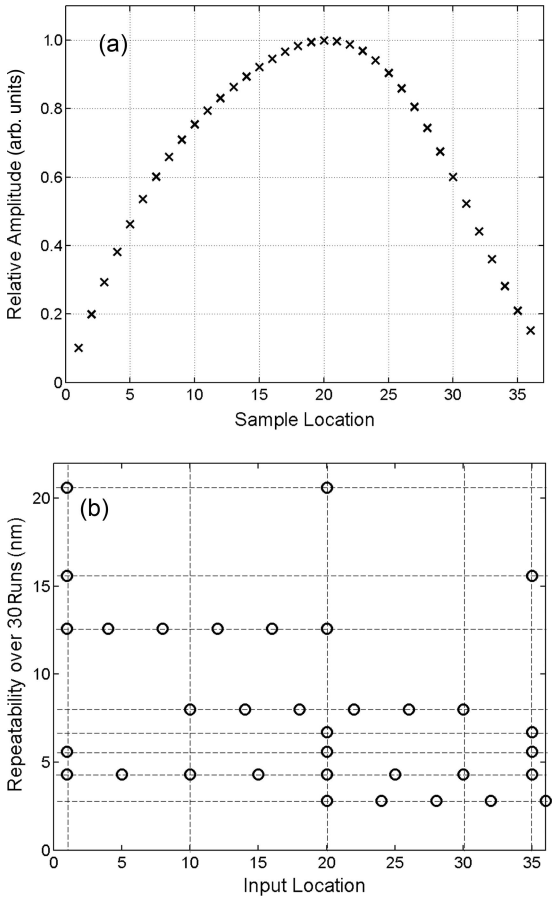

Fig. 8. Effects of the locations of the input patterns to the ANN, (a) possible input points for the ANN; and (b) measurement repeatability using different combinations of input points.

Table 6. Measurement Repeatability for Different Input Patterns to the ANN

Pattern Type	Input Locations	Standard Deviation in nm
1	1 20	20.6
2	1 35	15.6
3	1 4 8 12 16 20	12.6
4	10 14 18 22 26 30	8.0
5	20 35	6.7
6	1 20 35	5.6
7	1 5 10 15 20 25 30 35	4.3
8	20 24 28 32 36	2.8

the system. The results shown above were obtained using eight equally spaced sample points as the network inputs, and they are points 1, 6, 11, . . . 36. To investigate the influence of the locations of the input points, we trained a series of networks with different patterns of the input locations. Eight patterns were tested, and the results are shown in Fig. 8(b), where the patterns are arranged in descending order of the error levels.

It is obvious from the figure that high-frequency components are essential for precise measurements of the samples. It is interesting to observe that the best result is obtained with all low-frequency components omitted from the input. However, for different types of measurements, different input formats will be required. For example, if the track height is also a variable, a uniformly spaced input may be more appropriate for the network. It is therefore important to keep in mind that one input format will not suit all possible applications. The results shown in Fig. 8 are summarized in Table 6.

E. Removal of Incorrect Targets

Unlike human brains, the ANNs used here do not possess any understanding of the problems that are being tackled. The ANNs learn to solve a problem through a set of rules imposed by the learning algorithm. The target values used in Fig. 1 are the track widths normalized to within a range of 0.2 to 0.8 (the range of the output node is −1:1). Once an ANN is trained, its measured values can be converted back to physically meaningful, and hopefully accurate, values.

Consider the situation when one or more of the target values are incorrect. Instead of being exposed to a training set that possesses a certain logical pattern, the ANN is presented with one that contains faulty information. Under this condition, the ANN will return a large error. Figure 9(a) is a simulated result using the data from the Si sample. The target value of sample 11 was deliberately increased by 5% before the training. The effect of this on the training outcome is obvious, and that sample returns a very large error. In addition, the overall performance of the network has degraded significantly. Such a dramatic effect can also be used to solve the problem. By inspecting the error profile of the training set, training samples that return errors beyond a certain value

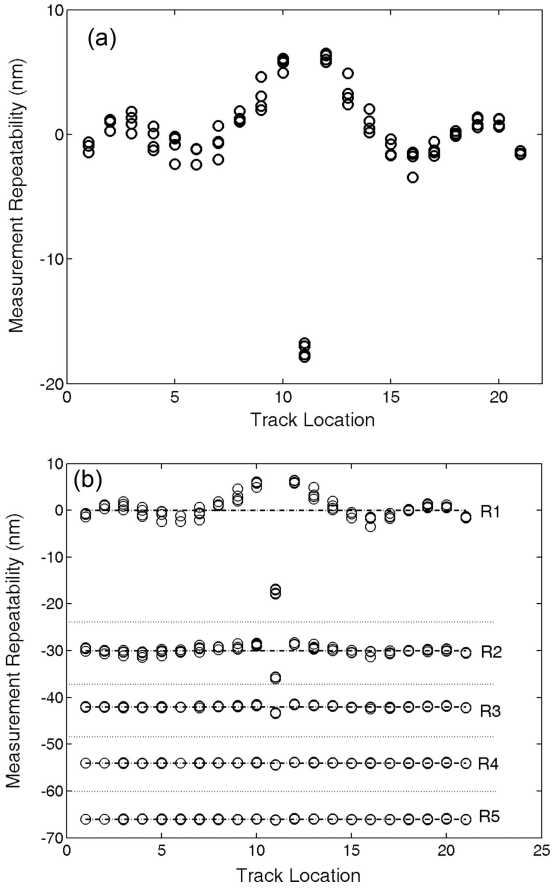

Fig. 9. Effects of incorrect target values: (a) measurement repeatability when a 5% error was imposed to one of the target value; and (b) the effects of the autocorrection routine on the measurement repeatability.

can be removed. The remainders are then used to train a second network. The situation is now similar to the missing tracks discussed above, and good quality networks can be expected.

Alternatively, those training samples with incorrect targets can be left with the other training samples, and a correction routine is used to reduce the error. The procedures of this correction routine are:

- Use the available data/target values to train a network as normal;
- Calculate the standard deviation error of the entire network, E_{net};
- Locate the sample S_x that has the largest standard deviation error, E_{sam};
- Adjust the target value of sample S_x by ΔE, where $\Delta E = E_{sam} - E_{net}$;
- Repeat training with the updated data set; and
- Repeat the process until the error is acceptable.

Essentially, the routine looks for any outliers in terms of errors, and adjusts their values until the errors for all individual samples are within a certain bound, which can be determined through SNR consideration of the system. Figure 9(b) shows the application of the routine to the data used for Fig. 9(a).

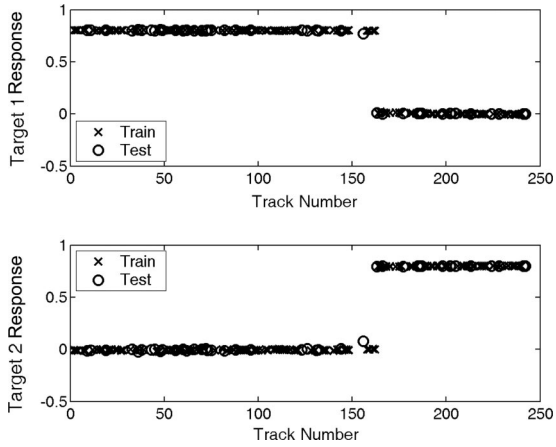

Fig. 10. Classification of single and double tracks.

The first row, R1, is the same as Fig. 9(a). Four iterations of the routine have been applied, with R2–R5 being the outcomes. The baselines of the four have been shifted for clarity.

It is important to point out that the autocorrection routine does not provide a more accurate network compared to simply removing the incorrect samples from the training. A possible advantage of the autocorrection routine is that it may yield networks with acceptable performance when the number of incorrect samples is large. This is an area that requires further investigation.

F. Classification

The results presented so far have shown the abilities of the ANNs in measuring very small dimensions to a very high degree of precision. For the technique to work properly, however, we need certain knowledge of the samples so that they can be presented to the appropriate ANNs. This knowledge may be available in some specialized areas. If, however, the system can determine the nature of the object without depending on prior information, the technique can become very powerful indeed. In this section, we will present results showing a possible way of tackling this problem.

ANNs may be used as classifiers. A well-known example is the unsupervised self-organizing network. We have tested such networks and they produced adequate results. We have found, however, that the multilayer perceptron networks used in this project so far are particularly suitable for this purpose. A network, identical to the previous ones except for having two outputs, has been trained to distinguish between single and double tracks of various dimensions. The output would return a value of 0.8 if a double track is presented to the input, and 0.2 for a single track. Using the chromium sample specified in Table 1, we have trained one such network, and the results are given in Fig. 10. As can be seen, the network yields perfect classification. Through simulations, we have found that ANNs work well even when the samples are very similar to each other. For example, it could distinguish a double track of 500 nm width (each track) and separation of 50 nm, to a single track of 1050 nm width. We have also successfully trained networks to separate samples of different structures. Since this aspect of the system is extremely important, it will be discussed further in the next section.

4. Future Work and Conclusion

The new results presented in this paper build on the previously reported results [1]. They demonstrate the measurement capability, the robustness, and the tolerance of the system.

With the current setup, the system can measure track width down to 60 nm with repeatability better than 2 nm. By optimizing the optical arrangement, a track width of less than 10 nm is well within the capability of the system. To achieve this level of performance, it is essential that the optical system should have high SNR and high stability. Typically, the system has an SNR in signal amplitude of around 3000:1. This is equivalent to a 9×10^9 detected photons per data point. For a scanning optical instrument, this level of SNR is readily achievable. Regarding stability, the system has a short-term drift over the length of a measurement cycle of less than 1 part in 5000. Again, with a proper isolated system, this stability can be improved further.

One rather surprising aspect revealed by the results is that, for such precise measurements, the requirement on the training set is relatively nondemanding. A good example is the coverage of the training sample set over the measurement range. Although it is not easy to determine exactly the most appropriate coverage, results in Table 4 show that there is a high degree of tolerance. Even if the training targets contain errors, by using the autocorrection technique described above, their effects are minimized. The one weakness of the system is that for an out of range profile, the error it produces is enormous. However, there are methods that can be implemented to overcome this problem, and they will be discussed further later in this section.

While the results presented here demonstrate the potential of the technique, further work is required before the technique can be extended to provide traceable metrological results and also applied to samples of different structures.

For metrological measurements the system will require a standard with which it can be calibrated. The standard will be measured by a high resolution system, such as an AFM or SEM. Training the network attempts to match the optical profiles with these measurements. Hence it could be argued that the optical/neural net system is being trained to make a "virtual" AFM (or SEM). To address the intrinsic uncertainty of the optical system, several factors have to be taken into account, such as the overall mechanical, electrical, and software stability of the system. Consequently, the uncertainty of the optical measurements will always be greater than the uncertainty of the AFM. For many applications, however, this in-

crease in the uncertainty is compensated for by the many virtues of an optical system.

The overall system must be sufficiently robust not only to provide valid results but also to recognize when it is unable to provide a valid result. The idea of calculating a confidence factor with each result has already been described. This would indicate to the user whether a result is suspect.

A limitation to the results that can be presented here is the lack of suitable samples for training and validating the system. In the ideal case a series of finely spaced linewidth structures is required to demonstrate not only the overall resolution of the system, i.e., that sub-100 nm structures can be resolved, but also a series of sub-100 nm linewidth structures with an increment of a few nanometers is required to demonstrate the overall performance of the system. Furthermore, for training the networks, a series of measurements is made at different points along a single linewidth structure. This does not take into account any variation in the width of the length of the linewidth structure, although hopefully this will be very small.

In the ideal case the system would be capable of measuring several different parameters related to the linewidth structure (e.g., sidewall angle and roughness). The potential of the system for measuring the width of double tracks also has been shown, and development of this will lead to the possibility of measuring periodic samples where the width and spacing of the structures are important.

The key to expanding the application area of the technique to features other than track widths and separations is being able to recognize the feature types of the profile. We have conducted a series of computer simulations on structures such as tracks with sidewall angles, or tracks spaced in close proximity to each other. It is clear that very precise measurements can be made provided appropriate ANNs are used for the profiles concerned. This is true for the type as well as the range of the features, and is backed up from some of the experimental results shown here. Obviously, if we have prior knowledge of the sample, then matching profiles with ANNs is not a problem. For many applications, this information is indeed available. Failing that, some mechanism will be required to distinguish between different samples. We will briefly discuss our approach here. A detailed description will be given in future publications.

The method we are exploring can be described as match and sift. The profile produced by the optical system is first compared to a series of templates, which are obtained from standard samples. When necessary, the comparison can have coarse and fine levels. Once the sample type is determined, it can be divided into different ranges before an ANN is applied to produce the desired measurement. The matching and sifting can be performed by a number of different techniques, for example, principal component analysis and project pursuit.

We have shown experimental results demonstrating the ability of our system to provide extremely precise measurements. The demands on the optical system can be described as moderate. If prior knowledge of the sample is available, the capability of the system is limited only by the quality of the signal, both in terms of drift and noise level. If object information is available, we can attempt to classify the sample profile by using the methods described. From preliminary investigation, this is an entirely feasible approach to adopt. One drawback of the method is that the application area is relatively limited since successful operation depends on the availability of the standard templates. On the other hand, the ability of the system can grow, since new templates can be added to the system. Thus many more sample types with more complicated structures can be measured. Although initially the system will require an AFM or SEM to provide standard measurements, it can be foreseen that one day all necessary templates are built into the system and it will function on its own.

Appendix A

The training of the ANN requires exposing the network to a set of reference profiles. To produce the reference profiles, a set of samples of known dimensions is needed. The values of the samples should cover the measurement range of interest. The optical system is then applied to the samples, producing a set of reference profiles with a set of target dimensions. In cases of limited availability of the reference samples, a process known as jittering may be employed. This involves applying the optical system to the sample set three to four times. The resulting reference set thus contains identical profiles with slight variations caused by random noise and experimental conditions. The jittering process helps to generalize the training of the ANN when the number of training samples is small.

Once the training profiles are available, they are conditioned before being inputted to the ANN. The conditioning involves:

- Removing the dc from the profiles;
- Normalizing the profiles by equalizing the absolute areas of the curves;
- Differentiating the normalized profiles;
- Selecting N number of points (usually eight) from the spectra of the differential profiles, inside the spatial bandwidth of the optical system; and
- Using the magnitudes of the N spectral components as the inputs to the ANN.

This data conditioning process is critical for the working of the ANNs. Other conditioning methods have been tried and none produced satisfactory results.

The training algorithm of the ANN is the well-known back propagation method. The weights of the network are initially assigned random values. The data from the reference profiles are then presented to the ANN, and the outputs are compared to the target values. The error functions are back propagated

through the network, and the weights are adjusted to reduce the errors in a least mean squared manner. The same set of training profiles is presented to the ANN repeatedly until certain predefined criteria, which may be the magnitude of the error or rate of change of the error, are met and the process is ended.

We 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 Program for Length Measurement 2002–2005.

References

1. 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," Meas. Sci. Technol. **16,** 2397–2404 (2005).
2. J. Nunn, W. Mirande, H. Jacobsen, and N. Talene, *Challenges in the calibration of a photomask linewidth standard developed for the European Commission,* VDE-VDI Conference Proceedings: Mask Technology for Integrated Circuits and Microcomponents, 53–68 (1997).
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. S. J. Howard, "Method for continuing Fourier spectra given by the fast Fourier transform," J. Opt. Soc. Am. **71,** 95–98 (1981).
7. M. H. Hayes, *Statistical Digital Signal Processing and Modeling* (Wiley, 1996).
8. G. Toraldo di Francia, "Resolving power and information," J. Opt. Soc. Am. **45,** 497–501 (1955).
9. Although it can be shown readily that if a second profile is measured under a different experimental condition this will provide sufficient information to resolve the ambiguity.
10. H. Wolter, in *Progress in Optics,* E. Wolf, ed. (North-Holland, 1961), Vol. 1, Chap. 5.
11. E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis,* 4th ed. (Cambridge University Press, 1927).
12. W. Lukosz, "Optical systems with resolving powers exceeding the classical limit," J. Opt. Soc. Am. **56,** 1463–1472 (1966).
13. W. Lukosz, "Optical systems with resolving powers exceeding the classical limit II," J. Opt. Soc. Am. **57,** 932–941 (1967).
14. 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).
15. D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty I," Bell Syst. Tech. J. **40,** 43–63 (1961).
16. H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty II," Bell Syst. Tech. J. **40,** 65–84 (1961).
17. H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty III: the dimension of the space of essentially time- and band-limited signals," Bell Syst. Tech. J. **41,** 1295–1336 (1962).
18. C. K. Rushforth and R. W. Harris, "Restoration, resolution, and noise," J. Opt. Soc. Am. **58,** 539–545 (1968).
19. K. Minami, S. Kawata, and S. Minami, "Surper-resolution of Fourier transform spectra by autoregressive model fitting with singular value decomposition," Appl. Opt. **24,** 162–166 (1985).
20. S. Haykin, *Neural Networks—A Comprehensive Foundation,* 2nd ed. (Prentice–Hall, 1999).
21. K. Gurney, *An Introduction to Neural Networks* (UCL Press, 1996).
22. N. B. E. Sawyer, C. W. See, M. Clark, M. G. Somekh, and J. Y. L. Goh, "Ultrastable absolute-phase common-path optical profiler based on computer-generated holography," Appl. Opt. **37,** 6716–6720 (1998).