Sub 0.1μm Track Width Measurement Using a Common Path Optical Interferometer and Artificial Neural Network.

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

[*]University of Nottingham, School of Electrical and Electronic Engineering, University Park, Nottingham, NG7 2RD, UK.

[^]Centre for Basic Thermal and Length Metrology. National Physical Laboratory, Queens Road, Teddington, Middlesex, UK, TW11 0LW

ABSTRACT

In this paper, we will describe a technique that combines a common path scanning optical interferometer with artificial neural networks (ANN), to perform track width measurements that are significantly beyond the capability of conventional optical systems.

Artificial neural networks have been used for many different applications. In the present case, ANNs are trained using profiles of known samples obtained from the scanning interferometer. They are then applied to tracks that have not previously been exposed to the networks. This paper will discuss the impacts of various ANN configurations, and the processing of the input signal on the training of the network.

The profiles of the samples, which are used as the inputs to the ANNs, are obtained with a common path scanning optical interferometer. It provides extremely repeatable measurements, with very high signal to noise ratio, both are essential for the working of the ANNs. The characteristics of the system will be described.

A number of samples with line widths ranging from 60nm-3μm have been measured to test the system. The system can measure line widths down to 60nm with a standard deviation of 3nm using optical wavelength of 633nm and a system numerical aperture of 0.3. These results will be presented in detail along with a discussion of the potential of this technique.

Keywords: Artificial Neural Network, interferometer, line-width measurement

INTRODUCTION

Providing calibrated line width standards for industry is an important measurement service. Over recent years as the feature size of these calibrated samples have decreased it has become impossible to use conventional optical microscopes for these measurements. Therefore there has been a move to using other, non-optical techniques, to achieve the necessary measurement resolution. These techniques include atomic force microscopy (AFM) and scanning electron microscopy (SEM). While these systems have lateral resolutions not matched by optical systems, they are not without problems of their own. They are expensive, difficult and time consuming to operate. The sample may also be damaged unless great care is taken when operating the system.

In this paper we will describe a technique that combines an optical interferometer with an artificial neural network, which is capable of measuring line widths substantially below 0.1 μm.

[1] eexrjs@nottingham.ac.uk, Tel +44 (0) 115 9515605, fax +44 (0) 115 951 5616, www.eee.nottingham.ac.uk

The factors that govern the lateral resolution of an optical microscope are well known: for a particular configuration, the bandwidth of the microscope is determined by the wavelength of the light used and the numerical aperture of the objective lens. Features smaller than a certain size will scatter most of the illumination light outside the aperture of the system, thus making it impossible to resolve the features. In addition, as the feature sizes of the object decrease, the optical profiles will no longer be linearly related to the actual object, and critical measurement of the dimensions will increasingly be affected by random noise associated with the system. One such example is shown in figure 1, where the FWHMs of the simulated, noiseless, track profiles are plotted against the actual widths. The diameter of the psf of the simulated microscope is 2.6 μm. It is apparent that when the track width goes below 1.5 μm, accurate measurement will become difficult.

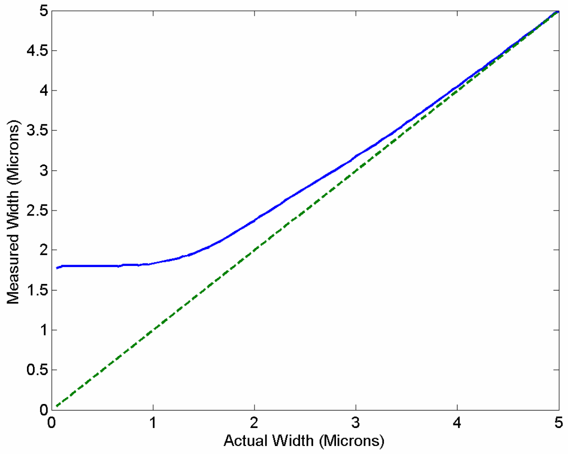

Figure 1: Actual Vs Measured Trackwidth for Simulated Data

There has been much research in the area of super-resolution, [1 2] with the aim of reconstructing the spatial frequency components of the object that are originally outside the system bandwidth. Many of the super-resolution techniques are based on the principle of analytic continuation [3 4] . To achieve any reconstruction of the object spectrum, the signal to noise ratio of the system needs to be extremely high, as discussed in the article by Cox and Sheppard [5] . In addition, the transfer function of the system needs to be known very accurately. In practice, therefore, the amount of improvement is limited.

We have taken a different approach, which is more suited for our application, where only an accurate value of the track width of the object is required. The information contained within one single parameter is much less than that contained in an extended spectrum, however small the latter may be [5] . The approach we have adopted makes use of an ultra-stable scanning interferometer, which is capable of producing sample profiles of high signal to noise ratio. We then process the profiles by using artificial neural networks, resulting in accurate measurements of line widths of tracks much smaller than those afforded by conventional optical microscopes. It should be reiterated that, at the moment, the technique only produces one single measurement and does not represent super-resolution, although work is underway to extract other parameters associated with the sample dimensions.

In the next section we will describe the optical system and the requirements demanded on it by the ANN. The characteristics of the ANNs suitable for our application will then be discussed. Experimental results, showing the capability of the technique, will be presented. Finally, areas of future work will be considered.

OPTICAL SYSTEM

The requirements of the signal processing technique on the optical system mean that it must be very stable, produce repeatable measurements and have a high signal to noise ratio. As the samples of interest will usually be purely phase objects the system must be sensitive to phase changes. A scanning interferometer is preferred as it meets all of the criteria.

The system used to obtain surface profiles is an ultra stable common path scanning optical interferometer[6].

The system employs a computer generated holographic (CGH) element as a beam splitter. The CGH creates two output beams from an incident collimated beam as shown in figure 2. The first beam is the unaltered zero order beam which is focused onto the sample surface by the objective lens. The second beam is focused to the back focal plane of the objective and is then collimated onto the sample surface by the objective lens. Upon reflection from the sample, the two beams traverse through the system and interfere to form parallel fringes. The fringe spacing is determined by the lateral offset of the hologram with respect to the optical axis, which alters the angle of the reference beam in the system.

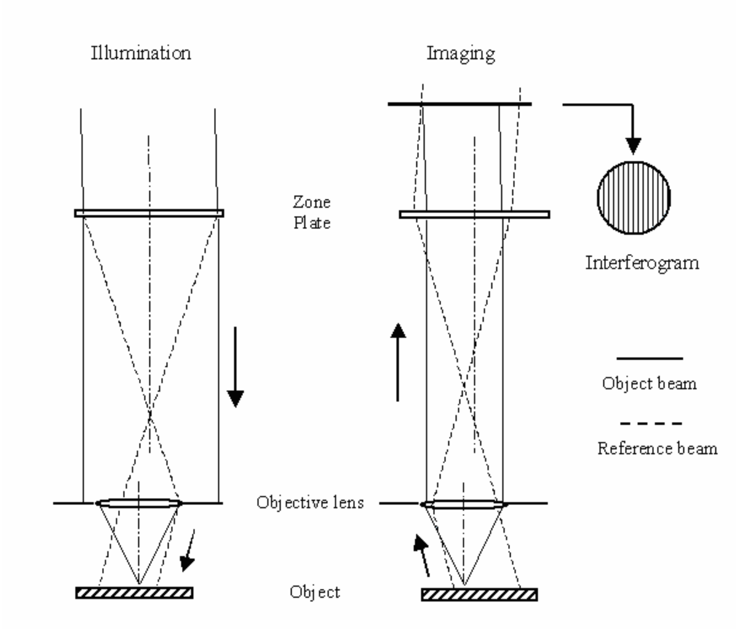

Figure 2: Optical arrangement of hologram and objective

As the object is scanned, local variations in surface height change the phase of the focused beam, whereas the average phase of the reference beam will essentially remain unchanged. Amplitude and phase profiles of the object are built up by recording the complex amplitude of the spectral component of the fringes at each scan location, by taking the Fourier transform of the fringe pattern. Figure 3 shows a typical fringe pattern recorded at the CCD camera, and its Fourier transform.

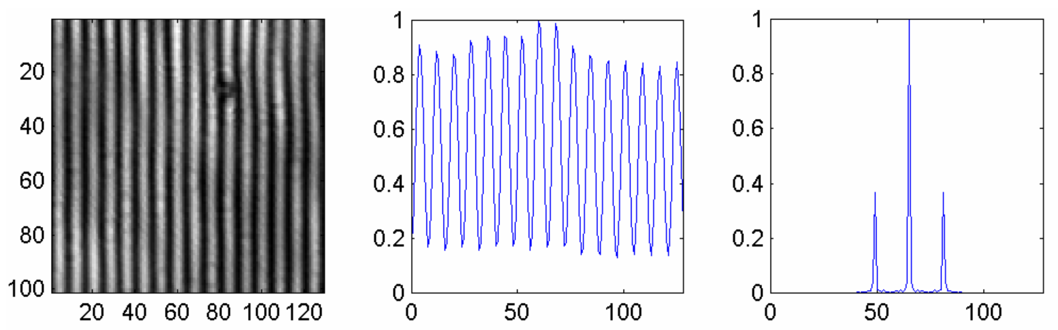

Figure 3: Fringes Captured at CCD, Line though fringes, Fourier transform of fringes

The common path nature of the system ensures that the effects of microphonics will be greatly reduced increasing the stability of the system and allowing operation close to fundamental limits. Both are critical considerations for the application of this project

ANN

Artificial neural networks are made up of many interconnected nodes. The nodes are simple computational units inspired by the biological neuron. Each neuron/node calculates the weighted sum of its inputs; this is the input into an activation function, the output of which forms the input to the next node. The topography of the network makes it a powerful computational device that has many applications. [7]

We use a feed forward network, which consists of an input layer a hidden layer and an output layer. Each layer is fully connected to the preceding layer as shown in figure 4. When in normal operation the network feeds the inputs to the hidden layer, the outputs from the hidden layer become the inputs to the output layer. The output from the final layer, for our application, is related to the trackwidth of the input object.

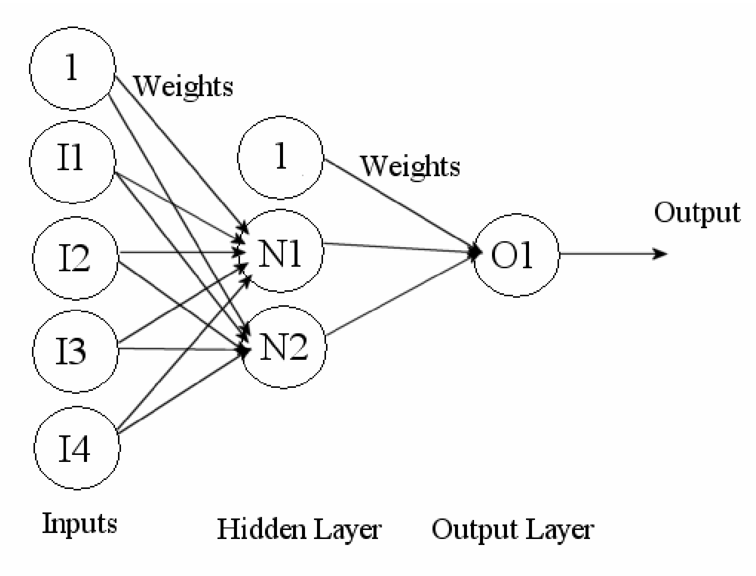

Figure 4: Schematic of Artificial Neural Network

Our network typically uses 8 inputs, 5 hidden nodes and 1 output node. The activation functions are hyperbolic tan functions. The network is trained using a training rule based on the Levenberg/Marquart[8] method as this produces fast reliable training.

The topography of the network was obtained by trying different network configurations with both simulated and experimental data and choosing a structure that produced reliable and repeatable results. As the network is only calculating one parameter the number of hidden nodes should be some where between the number of input and output nodes[9]. There is only one output node corresponding to the track width. Using eight inputs produced the best results for our data, too few nodes and the network was unable to train well and the errors were high, increasing the number of inputs not only slowed down the training process considerably but also did not improve and often degraded the training. Similarly using too few hidden nodes meant the network error was high and training was unreliable. Using too many hidden nodes again made the network difficult and time consuming to train and in some cases performed less well. The final selection of 8-5-1 took all of this into consideration and work equally well on both simulated and experimental data although similar topographies also work well.

The ANN is trained using a set of training data derived from the optical profiles and their *known* width values as training targets. The input patterns are presented to the network and an error value is calculated. This is back propagated through the network to update the weight values so that the overall error decreases. When the error no longer decreases the network finishes training and can be used to calculate the track width of previously unseen tracks.

The input patterns presented to the network are obtained from the optical profiles produced by the optical system after some processing. The profiles are first differentiated then their Fourier transforms taken. Then the spectral components in the pass band are sampled at 8 equally spaced locations and are used as the inputs to the ANN. The targets for each profile are scaled to fit into the output range of the output layer activation function.

The format of the input data is very important. Several different schemes were tried such as using the optical profile directly and using the Fourier transform of the optical profile both produced poor training results. Finally differentiating the data was tried as this would suppress the low spatial frequency components and enhance the high frequency components. The high spatial frequency components have proved to be the most important for changes in shape of the tracks as the track width reduces. This produced good training results, essentially because a lot of unimportant information was removed in the differentiation process.

Due to the limited number of distinct tracks on available samples, an early stopping technique and jittering [10] are used to help the network remain general. The network monitors the error in the testing set during training and training is stopped if the error of the testing set increases. If the error increases in the testing set but continues to decrease in the training set then the network has started to over train and memorise the training patterns.

RESULTS

A silicon sample comprising of 23 tracks in the range of 60-480nm has been measured. The tracks were etched in a silicon substrate and the height of each track was 45 nm. The tracks were separated by 60 microns. Each track was measured 4 times to build up a set of training patterns for the ANN. The numerical aperture of the optical system was 0.3 and the wavelength used was 633nm.

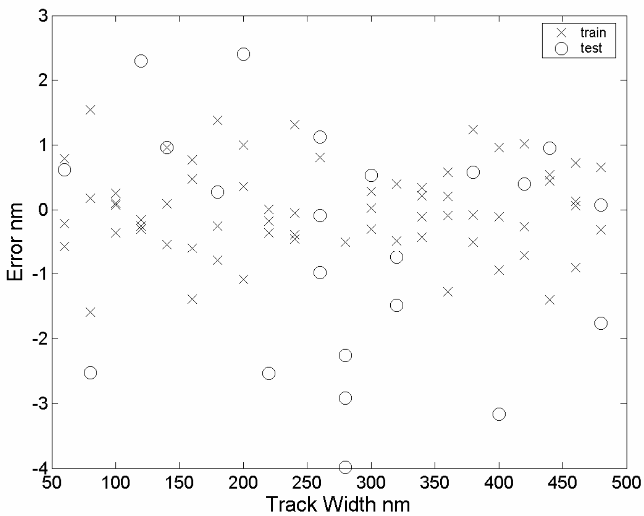

Figure 5: Experimental results

All of the available tracks on this sample were measured and divided into training and testing sets. 75% were picked at random to form the training set and the remaining used for the testing set. All of the tracks were processed as described in the previous section. Figure 5 shows the departure from the target value and the network output, the standard deviation across the range of 60-480nm is 2.5nm. The smallest track measured (60nm) is 42 times smaller than the point spread function of this system, which in this case was 2.6microns.

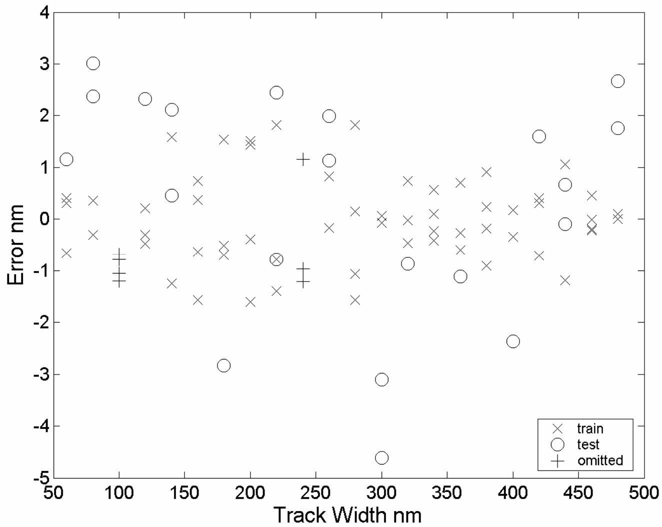

Figure 6: Experimental results showing generalized nature of network

To confirm the general nature of the trained network, training was repeated but with two tracks removed from the training process. Upon completion of the training, the network was applied to the two tracks originally omitted. Figure 6 shows the differences between the network response and the target value for the training, testing and the omitted tracks. The results for the omitted tracks are similar for the rest of the input tracks showing that the network is general.

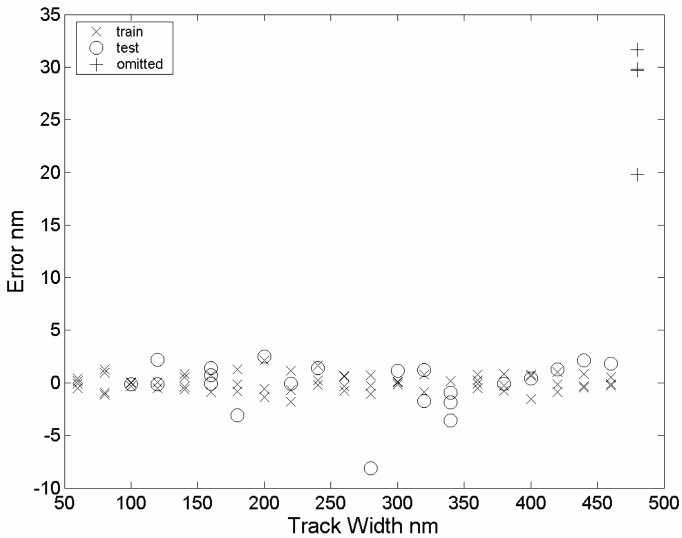

Figure 7: Experimental results showing out of range response

Another network was trained where the last track was left out of the training process. This illustrates the point that the network can only provide accurate values of the track width for the range of tracks it was trained on. Appling data for a track width outside of this range will produce a dramatic increase in the difference between the network output and target value. This is clearly shown in figure 7 where the deviation from the target value increases from around 3nm to an around of 30nm for the out of range track.

Figure 8 shows the effect of training with fewer input points. In this example the input data contained three points, one from the low spatial frequency end, one from the middle and one from the high frequency end. These were used to train an ANN and the network output deviations from the targets values are plotted below. The standard deviation for this example is 5nm. This has increased by a factor of two over the 8 input training case

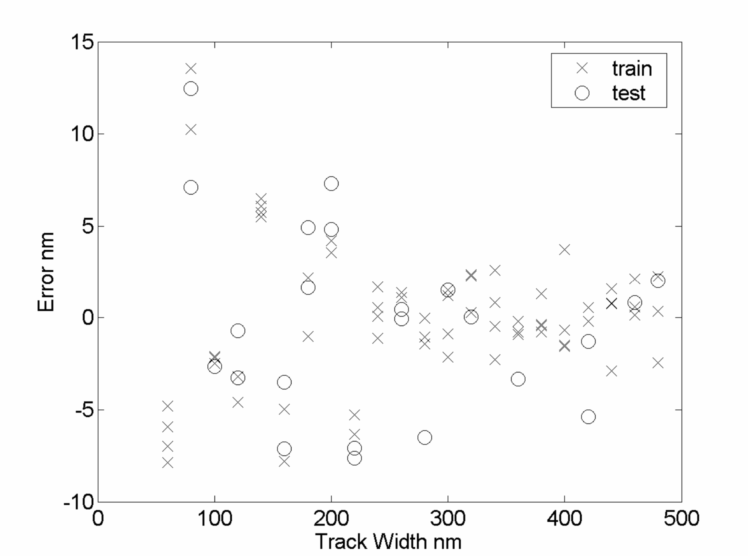

Figure 8: Training with fewer inputs

Another network was trained where the high frequency point from the set of three input points in the above example had been removed. This greatly reduced the performance of the network and the standard deviation of difference between the network response and the targets increases to over 10nm. This increase is due to two effects, firstly training with less input points and secondly the high spatial frequencies are very important for producing good results as the high frequencies are where the effects of reducing the track width are most prominent. More work is being carried out on the importance of the spatial frequencies and the number of inputs used to train the ANN.

The limit of this technique has yet to be met, as the 60nm track is the smallest feature we have available to measure. By increasing the objective NA and using a shorter wavelength we believe it should be possible to measure track widths down to 10nm.

Due to the success of this system when dealing with single tracks, we have considered extending its use to multiple track objects. Computer simulations of this shows that this technique is appropriate for obtaining both the track width and separation for a double track structure, experiments are underway to confirm this.

The possibility of a more suitable optical system is being considered with the aim of providing the differentiation optically.

MORE HERE?

CONCLUSIONS

A combined system of an ultra stable common path interferometer and artificial neural network for line width measurement has been presented. Tracks widths down to 60 nm have been measured with a numerical aperture of 0.3 and optical wavelength of 633nm.

This technique has proven to be very powerful in extending the capability of the optical system for the application of track width measurement. The solution obtained by the ANN is general in that any other widths in the trained range produce similar error levels. However for this system to be useful for providing *standard* measurements work regarding the uncertainty of the optical system and the ANN needs to be carried out. This is not trivial with regards to the ANN as discovering what the ANN is actually doing is not straightforward.

ACKNOWLEDGEMENTS

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

REFERENCES

1. Toraldo Di Francia, G., "Resolving power and information," J. Opt. Soc. Am. **45**, 497 – 501 (1955).

2. Harris, J.L., "Diffraction and resolving power," J. Opt. Soc. Am. **54**, 931 – 936 (1964).

3. Whittaker, E.T., A Course of Modern Analysis: An Introduction to The General Theory of Infinite Processes and of Analytical Functions, (Cambridge University Press, Cambridge, 1902).

4. Morse, P.M. and Feshbach, H., Methods of Theoretical Physics, Part I, (McGraw-Hill, New York, 1953).

5. Cox, I.J. and Sheppard, C.J.R., "Information capacity and resolution in an optical system," J. Opt. Soc. Am. **A 3**, 1152 – 1158 (1986).

6. Sawyer N, Chung Wah See, Clark M, Somekh MG, Goh JYL. Ultrastable absolute-phase common-path optical profiler based on computer generated holography. 1998, Appl Opt. vol 37(28), 6716-6720.

7. wide range of applications

8. Hagan, M.T. and Menhaj, M., "Training feedforward networks with the Marquardt algorithm," IEEE Trans Neural Networks, **5**, 989–993, (1994).

9. Sarle, W.S., "Stopped Training and Other Remedies for Overfitting," Proc. 27th Symposium: Interface of Computing Science and Statistics, 352-360, ftp://ftp.sas.com/pub/neural/inter95.ps.z (1995).

10. Blum, A. "Neural Networks in C++", NY: Wiley. p. 60. (1992)