# Automated Direct Patterned Wafer Inspection

Babak H. Khalaj Hamid K. Aghajan Thomas Kailath

Information Systems Laboratory
Department of Electrical Engineering
Stanford University
Stanford, CA 94305

## Abstract

*A self-reference technique is developed for detecting the location of defects in repeated pattern wafers and masks. The application area of the proposed method includes inspection of memory chips, shift registers, switch capacitors, and CCD arrays. Using high resolution spectral estimation algorithms, the proposed technique first extracts the period and structure of repeated patterns from the image to sub-pixel resolution, and then produces a defect-free reference image for making comparison with the actual image. Since the technique acquires all its needed information from a single image, there is no need for a database image, an scaling procedure, or any a-priori knowledge about the repetition period of the patterns.*

## 1 Introduction

The task of detection and localization of defects in VLSI wafers and masks is an essential but exhausting procedure. As the complexity of integrated circuits is increasing rapidly, the need to automate the inspection of photomasks and wafers becomes a more important necessity for maintaining high throughput and yield in the fabrication processes. Human visual inspection and electrical testing are the most widely used methods for defect detection; however, this is a time consuming and difficult task for people to do reliably. On the other hand the usage of electrical test is inherently limited to off-line and overall functional verification of the chip structure, and can only be accomplished after the fabrication is completed; it cannot be applied to on-line and layer by layer inspection of the wafer during the fabrication process. In addition to the need for inspecting wafers, the inspection of the mask pattern is critical because any defect on the mask is transferred to the wafers.

Typical patterns found in wafers and masks can be put into three main classes [1]:

- Constant areas
- Straight lines
- Repeating structures

The repeating structure class covers two different cases. The first includes repeated patterns within a single chip such as memory areas, shift registers, adders, and switch capacitors. The chips themselves considered as repeated patterns on a wafer can be included in the second case. As another potential example of repeating patterns, one can mention arrays of charge-coupled devices (CCDs) arising in imaging systems and cameras.

Most inspection techniques fall into one of the following general categories: methods for checking generic properties and design rules, and methods based on image-to-image comparison. In the first category, the image is tested against a set of design rules or local properties and violations are reported as defects. An example of this kind of techniques is the work of Ejiri *et al* [2] that uses an expansion-contraction method to locate the defects. In image to image comparison methods, the image taken from the wafer is compared either with an ideal image stored in a database, or with the image taken from another region of the same wafer that is supposedly identical to the image under the test. A fairly complete review of the related literature may be found in [3].

Several optical inspection techniques have been developed for locating and classifying defects on masks and wafers. In spatial filtering methods, the spectrum of the perfect image in the Fourier transform domain is filtered out from the image and an image that includes only defective patterns is obtained. This technique can be implemented through a holographic approach [4] to attenuate distortions; however, since it is difficult to



0-8186-2840-5/92 $03.00 © 1992 IEEE

filter out only the frequencies of the acceptable pattern, the signal to noise ratio of the resulting defect image is generally poor [5].

Most commercial inspection systems compare the chip patterns with a pre-stored image in a database. This requires a large volume of data as a reference. A data conversion step is also needed to make the scaling of the stored data equal to that of the inspected image [5]. As the size of devices decrease, proper adjustment of the scales for doing the required comparisons becomes more difficult to achieve. As an alternative to this approach that avoids the need for a large database, images of two adjacent dies can be compared; however, the detection is limited by step-and-repeat errors and also the errors in synchronizing the location of the two scanner beams over the die.

As was mentioned earlier, an important category of defect inspection applications is the inspection of repeated patterns on masks and wafers. This is a field of application for image-to-image comparison methods in which the repeatedness of the patterns is used. However, most existing systems that perform image to image comparison face the following difficulties. In order to compare images with each other, or with a reference database, accurate registration is necessary. This includes both problems of alignment and scaling and introduces a tradeoff between the minimum detectable defect size and the expense and throughput of the systems that compare the image with another image or a database image. Moreover, if an image to database comparison method is chosen, there is also the need to simulate the imaging and development processes in order to produce the database.

A self-reference technique that avoids the mentioned difficulties was developed by Dom *et al* [6], in which the comparison is made using the repeated cells in the image. In this method prior knowledge about the period of repetition is assumed and scaling of the image is adjusted accordingly; then each pixel is compared with two corresponding pixels in left and right neighboring patterns.

In this paper, we propose a technique for extracting the structure of repeated patterns, or the *building block*, from the acquired image itself; and then detecting the defects by comparing the resulting building block with the image. Extracting the repeated structure from the image eliminates the need for producing a database and avoids the scaling problem as well. There is also no need for the image to exactly contain a certain number of pattern blocks. Also rather than using only pixels from neighboring patterns, the proposed method exploits information contained in the whole image to decide whether a pixel corresponds to a defect or not. Moreover, the techniques used in the proposed approach yield sub-pixel resolution for estimating the size of the building block.

The organization of the rest of the paper is as follows. In Section 2, the details of the proposed technique are presented. Section 3 contains several examples of application of the proposed technique for detection of defects in real wafer images. Section 4 summarizes the results and offers some concluding remarks. In the proposed algorithm, the periods of the patterns are estimated by techniques of sensor array processing.

## 2 The Defect Detection Algorithm

The proposed defect detection algorithm is composed of three steps. In the first step the repetition periods of the patterns in both the horizontal and vertical directions are estimated by a high-resolution method. After obtaining good estimates of these periods, the building block of the image is extracted by a proper sub-pixel weighted sum of the repeated patterns throughout the image. In the final step the location of defects is determined by subtracting shifted versions of the building block from the image. In order to reduce the number of false alarms especially at the edges of the image, some ideas from fuzzy-logic and median filtering are used in this step. By passing the resulting difference image through a proper threshold, the location of the defects is obtained.

### 2.1 Estimating the Period of Repeatedness

The general problem of extracting the spectral components of a signal arises in a wide range of application areas such as in communications systems, geophysical data processing, and vibration analysis. The objective in time series analysis is to obtain estimates of the sinusoidal frequencies in a discrete time signal presumed to be composed of a superposition of sinusoids added with noise. In a more generic case, one can consider general exponential sinusoids with decaying or growing envelopes. Mathematically, if the $N \times 1$ vector $\mathbf{x}_N$ denotes the observed time series, and $\mathbf{\Lambda} = [\lambda_1, \ldots, \lambda_d]$ is the unknown vector of exponential terms, the observed signal can be expressed as

$$x_k = \sum_{i=1}^{d} s_i \lambda_i^{k-1} + n_k \qquad (1)$$

where $s_i$ contains the amplitude and phase of the $i^{th}$ exponential $\lambda_i = e^{j\omega_i}$, and $n$ is the additive noise term.

An efficient technique adopted from the field of sensor array processing, called ESPRIT is used to estimate the frequencies $\lambda_i$ [7, 8]. In our application, we are looking for estimates of the periods of a periodic image in the horizontal and vertical directions. Therefore, the problem is slightly different from the harmonic retrieval problem in that the goal here is to find only the fundamental frequency of the signals.

As a matter of background, there are a number of different ways for estimating the period of a periodic signal. Among these methods one can mention the FFT and the autocorrelation approaches. However, there are two major drawbacks for these methods. In our problem we are dealing with two-dimensional images and so the amount of computations will be very high with either of there methods. The second and more important disadvantage of these standard methods is their limited resolution, which depends on the width of the signal. In typical images of wafers, the number of periods of the pattern can be as low as 5 and so we want to estimate the period of a signal by only observing a small number of its periods. Since the resolution of the FFT method for a given signal frequency is proportional to the number of observed periods of the signal, the period that is estimated by this method in these situations is not acceptable. Since in general the period of the patterns is not an integer number of the pixels of the image, it is necessary to use a high resolution (subpixel) algorithm for period estimation. We shall present such a method below.

First let us note that the estimation of the horizontal and vertical periods of a periodic image are separable problems in nature and by transforming the problem into two one-dimensional problems, large savings in computational load can be achieved. One way of doing this transformation is described below.

The image is first projected along its horizontal and vertical axes and two one-dimensional vectors are produced. Then, the aforementioned high-resolution techniques of time series analysis can be applied to each of these vectors in order to obtain sub-pixel estimates of the horizontal and vertical periods of repeatedness in the image.

Let the image be denoted by the matrix $\mathbf{F}$. We only treat here the estimation of the horizontal period of the image; an identical procedure is used for estimating the vertical period. The rows of $\mathbf{F}$ are $1 \times N$ vectors and are named $\mathbf{f}_i$ , $i = 1, \ldots, N$, so that we can write

$$\mathbf{F} = \begin{bmatrix} \longleftarrow & \mathbf{f}_1 & \longrightarrow \\ & \vdots & \\ \longleftarrow & \mathbf{f}_N & \longrightarrow \end{bmatrix} \tag{2}$$

The simple horizontal projection of the image produces an $N \times 1$ vector $\mathbf{x}$ whose elements are

$$x_i = \sum_{j=1}^{N} (\mathbf{f}_i)_j \quad , \quad i = 1, \ldots, N \tag{3}$$

Then, the goal is to estimate the fundamental frequency of the signal $\mathbf{x}$. This is called the *harmonic retrieval* problem in spectral estimation literature. Harmonic retrieval problems arise in many applications where a time series vector is given that is assumed to be a superposition of several cisoids contaminated in noise and the goal is to extract the frequencies of the present cisoids. Several high resolution methods have been developed for estimating these frequencies [7, 8]. If the $N \times 1$ vector $\mathbf{x}$ denote an observed time series, and $\mathbf{\Lambda} = [\lambda_1, \ldots, \lambda_d]$ be the unknown vector of exponential terms. Then, the observed signal can be expressed as

$$x_k = \sum_{i=1}^{d} s_i \lambda_i^{k-1} + n_k \tag{4}$$

where $s_i$ contains the amplitude and phase of the $i^{th}$ exponential $\lambda_i = e^{j\omega_i}$, and $n$ is the additive noise term. In matrix formulation, the entire time series can be written as

$$\mathbf{x} = \begin{bmatrix} 1 & \cdots & 1 \\ \lambda_1 & \cdots & \lambda_d \\ \vdots & \ddots & \vdots \\ \lambda_1^{N-1} & \cdots & \lambda_d^{N-1} \end{bmatrix} \begin{bmatrix} s_1 \\ \vdots \\ s_d \end{bmatrix} + \begin{bmatrix} n_1 \\ \vdots \\ n_N \end{bmatrix} \tag{5}$$

For applying eigenstructure methods of sensor array processing we need to compute the sample covariance matrix of the measurements. The above formulation views the data as one snapshot of a uniform linear array. This defines only a one-dimensional signal subspace. To obtain a subspace of dimension $d$, we divide $\mathbf{x}$ into $P$ vectors of length $m$ by sliding a window of size $m$ over the data. The value of $m$ should satisfy $d < m \leq N - d + 1$. Then, $P$ plays the role of the number of snapshots here as compared with array processing formulation. Thus, the data is rearranged as follows

$$\mathbf{X}_m = \begin{bmatrix} x_1 & x_2 & \cdots & x_{N-m+1} \\ x_2 & x_3 & \cdots & x_{N-m+2} \\ \vdots & \ddots & \vdots & \vdots \\ x_m & x_{m+1} & \cdots & x_N \end{bmatrix} \tag{6}$$

$$= \mathbf{A}_m(\mathbf{\Lambda}) \begin{bmatrix} \mathbf{s} & \mathbf{\Phi s} & \cdots & \mathbf{\Phi}^{N-m}\mathbf{s} \end{bmatrix} + \begin{bmatrix} n_1 & n_2 & \cdots & n_{N-m+1} \\ n_2 & n_3 & \cdots & n_{N-m+2} \\ \vdots & \ddots & \vdots & \vdots \\ n_m & n_{m+1} & \cdots & n_N \end{bmatrix} \quad (7)$$

where

$$\mathbf{A}_m(\mathbf{\Lambda}) = \begin{bmatrix} 1 & \cdots & 1 \\ \lambda_1 & \cdots & \lambda_d \\ \vdots & \ddots & \vdots \\ \lambda_1^{m-1} & \cdots & \lambda_d^{m-1} \end{bmatrix} \quad (8)$$

and $\mathbf{\Phi} = \text{diag}\begin{bmatrix} \lambda_1 & \cdots & \lambda_d \end{bmatrix}$. The vector $\mathbf{s}$ contains the amplitudes of different frequency components and is not important in our analysis.

With the above formulation, the signal subspace techniques of sensor array processing can be applied to estimate the significant frequencies of the signal $\mathbf{x}_N$. A computationally efficient technique for estimating the frequency components in the above problem is the ESPRIT algorithm [7, 8]. ESPRIT assumes the availability of measurements from two identical subarrays that are displaced from each other by a displacement vector $\mathbf{\Delta}$. In the time series analysis, the measurements come from equally spaced time instants. So this problem possesses the structure required by ESPRIT and the frequencies of the sinusoids in the time series signal can be estimated by this method.

In our application, the signal $\mathbf{x}_N$ is a periodic signal and has a set of harmonically related frequency components. In other words, we only need to estimate the frequency of the first harmonic of the signal. If the signal obtained by projection has a large component in its principle frequency (which is true for square waveforms), then by choosing the number of components equal to 4 (corresponding to the first and second harmonics in positive and negative frequencies), one can directly obtain the period of the signal using a subspace method. However, in general, there may be a relatively large amount of energy in higher harmonics, and in these cases one can estimate the value of these higher harmonics of the signal and since the frequencies of these harmonics should be multiples of the principle frequency, this frequency can be extracted by a simple linear least-squares method.

As mentioned before one of the most important characteristics of these subspace fitting methods is their ability to provide high resolution estimates of the parameters. In our application, this leads to subpixel resolution in estimating the periods. The estimated periods of row and column projections determine the size of the building block of the image. The procedure of extracting the building block is described in the next section.

## 2.2 Extracting the Building Block

Given estimates of the period of the patterns in both horizontal and vertical directions, the building block is constructed by simply shifting a window of proper size through the image and adding the corresponding pixel values together. Since in general the size of the building block is not an integer number of pixels, it is necessary to shift the window by subpixel values and use interpolation to find the values of subpixel points. A simple linear interpolation proves to be sufficient and is used for this purpose.

If the estimated horizontal and vertical periods of the image are $T_x$ and $T_y$ respectively, the size of the building block $\mathbf{BB}(k,l)$ will be

$$\begin{aligned} 1 \leq \quad & k \quad \leq int(T_x) + 1 \\ 1 \leq \quad & l \quad \leq int(T_y) + 1 \end{aligned}$$

Denoting the value of the $N \times N$ image at location $(k,l)$ by $\mathbf{F}(k,l)$, the following equations are used to construct the building block

$$\begin{aligned} \mathbf{BB}(k,l) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} & (1-r_i)(1-s_j)\,\mathbf{F}(k_i+k, l_j+l) \\ & + r_i(1-s_j)\,\mathbf{F}(k_i+k+1, l_j+l) \\ & + (1-r_i)s_j\,\mathbf{F}(k_i+k, l_j+l+1) \\ & + r_i s_j\,\mathbf{F}(k_i+k+1, l_j+l+1) \quad (9) \end{aligned}$$

where

$$\begin{aligned} n_1 &= int(N/T_x) \\ n_2 &= int(N/T_y) \\ k_i &= int(T_x * i) \\ l_j &= int(T_y * j) \\ r_i &= T_x * i - k_i \\ s_j &= T_y * j - l_j \end{aligned}$$

By averaging among all of the blocks in image, the amount of noise and the effect of the defects are reduced considerably and a good estimate of building block is obtained in this way. It should also be mentioned that if the sizes of defects in an image are so large that the computed building block is no longer a good estimate of the true value, then although the defect can not be localized exactly, the algorithm will still give an alarm and will reject the sample.

### 2.3 Detecting the Defects

When the building block is constructed, one might simply compare each point of the original image with the corresponding point in the building block, and if the difference is larger than a threshold, that point may be identified as a defect. However, because of the quantization effects at the edges of the image, in practice each point should also be compared with the eight neighboring points of its corresponding point in the building block. Having computed the difference between each point of the image and all 9 points in the neighborhood of its corresponding point in the building block, that point of the image is assigned a value equal to the minimum of the absolute value of these differences. It is this value that gives a measure of the probability that the point is a defect.

Mathematically, the above procedure can be stated by the relation

$$\min_{-1\leq i,j\leq 1} | \mathbf{F}(m,n) - \mathbf{BB}(k+i, l+j) | \qquad (10)$$

where

$$\begin{aligned} k &= m - int(m/T_x) * T_x \\ l &= n - int(n/T_y) * T_y \end{aligned}$$

By transforming all the points of the image to such a difference image, and using a proper threshold value, one can classify the points of image as either defects or nondefects. The value of the threshold generally depends on the contrast of the image and the amount of the difference in intensity which is supposed to be interpreted as a defect and can be chosen accordingly.

## 3 Experimental Results

In this section we apply the developed techniques to the detection of defects in images that have repeated pattern blocks.

In these examples, real images taken by an optical microscope are considered. In this case the period in either of the two directions is not an exact number of pixels and the sub-pixel algorithm is necessary to obtain a good estimate of these non-integer numbers. In Figs. 1 and 2 the image and its projection along the horizontal axis are shown. In fact the signal in Fig. 2 is the signal whose period is to be estimated. The spatial frequencies of this signal as estimated by the ESPRIT algorithm are shown in Fig. 3. Finally the estimated building block of this image and the location of the defects are shown in Figs. 4 and 5 respectively.

Fig. 6 shows the image of another patterned wafer, and Figs. 7 and 8 show the extracted building block and the location of the defects in this example.

The third example considers another real image, which consists of two interlaced subimages that are shifted with respect to each other (Fig. 9). As a result, there are some rows of the image that are similar but shifted versions of each other (Fig. 10), and by simple projection of the image along the horizontal axis we obtain a signal that has almost twice the true frequency of the patterns. In such cases by a proper preprocessing of the image (see below), we obtain another signal that has the correct pattern frequency (Fig. 12). By estimating the period of this signal, the building block and the location of defects are obtained accordingly (Figs. 13 and 14). To handle these cases a simple preprocessing step can be applied to the image as follows. First each row is *centered* by removing its mean. Let the new rows be denoted by $\mathbf{z}_i$ , $i = 1, \ldots, N$. The average vector of all rows is calculated and called $\mathbf{a}$. Then each row of the image is projected onto the plane orthogonal to $\mathbf{a}$. By this transformation the two vectors that were shifted versions of each other will be transformed to vectors 180 degrees apart. In Fig. 11, a simple diagram is shown to visualize this situation. Vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ represent the two similar and shifted rows and their projections onto the space orthogonal to $\mathbf{a}$ are along directions $\mathbf{q}_1$ and $\mathbf{q}_2$, respectively. After this preprocessing step, the inner product of each of the two kinds of rows with a selected row vector will have opposite signs to each other. An example of such cases is presented in Fig. 9. In Fig. 10(a) and (b), the two similar and shifted rows of the image are shown. Simple projection of these rows results in similar values in the projection vector $\mathbf{x}$ which is shown in Fig. 12(**a**). However, by implementing the preprocessing step and multiplying the transformed image by one of these typical vectors, the two shifted vectors will be transformed to two opposite numbers and so the period of the projection will be equal to the true period of the patterns. In the above example, the resulting projection vector $\mathbf{x}$ has the form shown in Fig. 12(**b**) in which the effects of similar rows have been converted to opposite values.

## 4 Conclusion

An automatic self-reference technique has been developed for detecting defects on repeated pattern wafers and masks. The technique uses information from the image to extract the building block of the repeated structure and locates the defects by subtract-

ing a regenerated defect-free image from the acquired image. There is no need in this method for a reference database image or a scaling procedure. Potential application areas for the proposed method include inspection of memory chips, shift registers, switch capacitors, and CCD arrays. Since the proposed algorithm extracts the required information from a single image, it may be applied for automatic detection of any defect or non-regularities in a repeating two-dimensional signal. In the current paper its application to images from the area of wafer and mask defect inspection has been considered, but the ideas of the proposed approach may also be extended to other areas that deal with repeated structues such as crystallography.

## Acknowledgements

This work was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under contract F49620-90-C-0014 (the United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon). The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

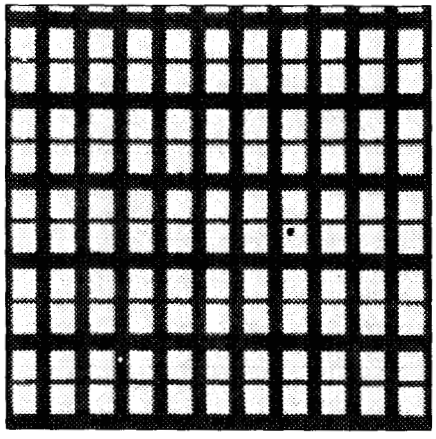

Figure 1: An image from a repeated pattern chip.

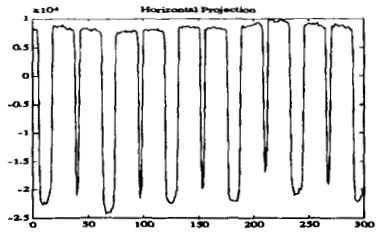


Figure 2: Projection vector along the horizontal axis of the image in Fig. 1.

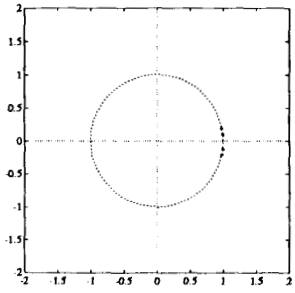

Figure 3: Horizontal spatial frequencies of the image in Fig. 1 estimated by the ESPRIT algorithm.

Figure 4: The building block of the image in Fig. 1.

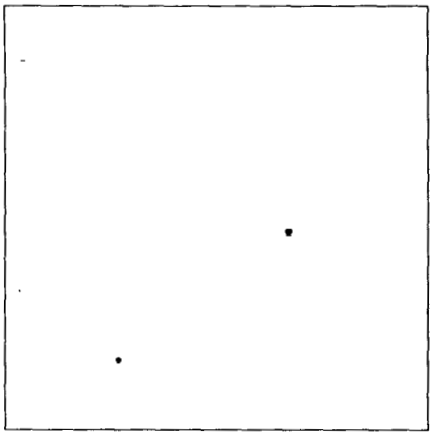

Figure 5: Defects of the image in Fig. 1 extracted by the proposed technique.

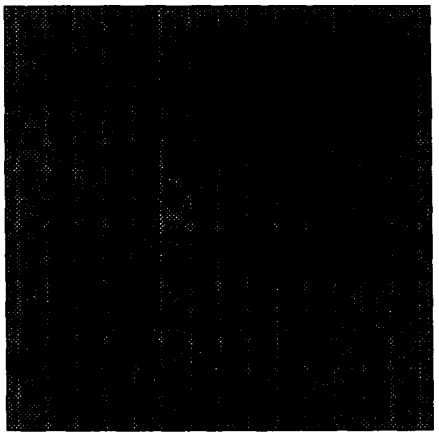

Figure 6: Another image from a repeated pattern chip.

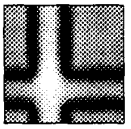

Figure 7: The building block of the image in Fig. 6.

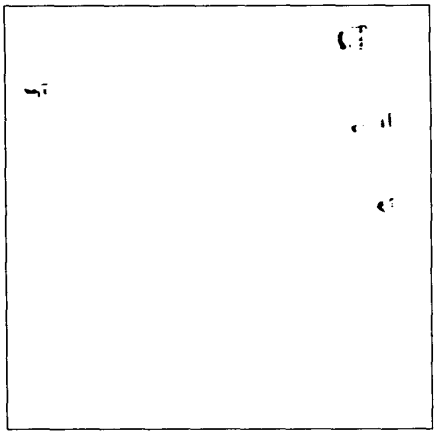

Figure 8: Defects of the image in Fig. 6 extracted by the proposed technique.

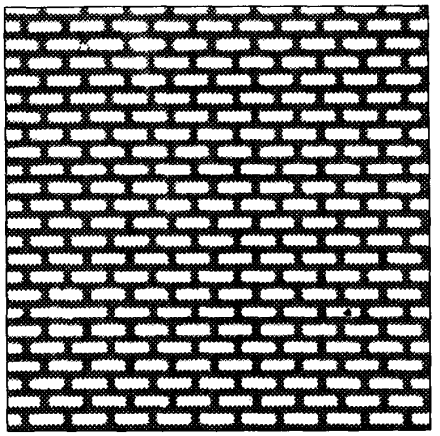

Figure 9: An image from a chip with similar and shifted rows.

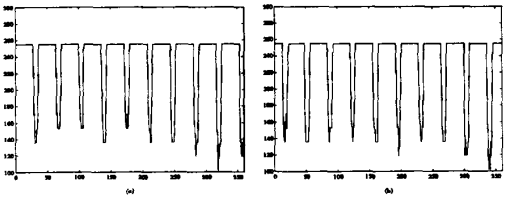

Figure 10: Plots of two rows of the image of Fig. 9 that are similar but shifted versions of each other.

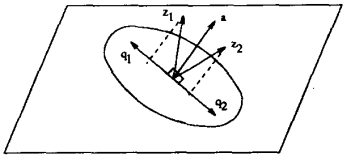


Figure 11: A diagram showing how two similar and shifted rows of the image are mapped to vectors with opposite directions.

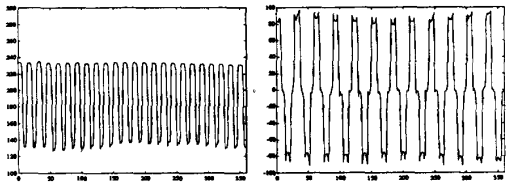

Figure 12: (**a**) Result of simple projection of the image of Fig. 9 along horizontal axis. (**b**) Result of applying the projection after preprocessing the image (see text).

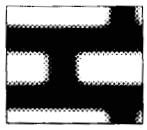

Figure 13: The building block of the image in Fig. 9.

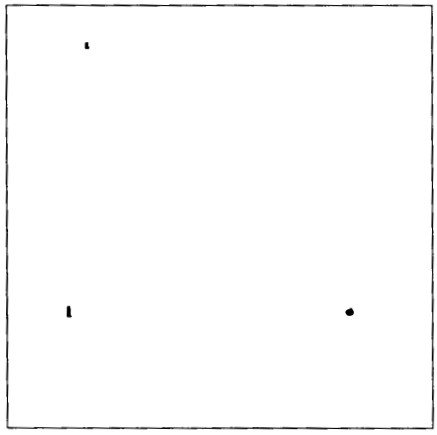

Figure 14: Defects of the image in Fig. 9 extracted by the proposed technique.

## References


[1] F. Babian, *Optical Defect Detection Limits in Semiconductor Wafers and Masks*, PhD thesis, Stanford University, Stanford, CA 94305, 1986.

[2] M. Ejiri, T. Uno, M. Mese, and S. Ikeda, "A Process for Detecting Defects in Complicated Patterns", *Computer Graphics and Image Processing*, **2**:326–339, 1973.

[3] A. E. Kayaalp and R. Jain, "A Knowledge Based Automatic On-Line Wafer (IC) Inspection System", In *Proc. of VISION 85*, pages 117–130, 1985.

[4] L. H. Lin, D. L. Cavan, R. B. Howe, and R. E. Graves, "A Holographic Photomask Defect Inspection System", In *Proc. of SPIE, Optical Microlithography IV*, pages 110–116, 1985.

[5] S. Chae, *Defect Detection and Classification in VLSI Pattern Inspection*, PhD thesis, Stanford University, Stanford, CA 94305, 1987.

[6] B. E. Dom, V. H. Brecher, R. Bonner, J. S. Batchelder, and R. S. Jaffe, "The P300: A System for Automatic Patterned Wafer Inspection", *Machine Vision and Applications*, **1**(3):205–221, 1988.

[7] A. Paulraj, R. Roy, and T. Kailath, "Estimation of Signal Parameters by Rotational Invariance Techniques (ESPRIT)", In *Proc. of 19th Asilomar Conference on Circuits, Systems and Comp.*, 1985.

[8] R. Roy and T. Kailath, "ESPRIT: Estimation of Signal Parameters via Rotational Invariance Techniques", *IEEE Trans. on ASSP*, **37**(7):984–995, July 1989.