STATISTICAL THRESHOLD METHOD FOR SEMICONDUCTOR INSPECTION

Akira Hamamatsu[1], Hisae Shibuya[1], Yoshimasa Oshima[1], Shunji Maeda[1],
Hidetoshi Nishiyama[2], Minori Noguchi[2]

[1]Hitachi, Ltd. Production Engineering Research Laboratory
[2]Hitachi High-Technologies Corporation, Nanotechnology Products Business Group

Abstract

Optical inspection technology aimed at high sensitivity inspection, which is expected to have applications in dark-field systems for semiconductor wafers, is proposed. A common method of wafer inspection is the die-to-die comparison method, where a subtraction image is compared to a threshold to find defects. An optimum threshold must be set in order to detect only defects and prevent misdetection of normal patterns with noise due to process variations. Various methods that involve segmenting regions and setting thresholds for each region have been reported, although difficulties remain. A statistical threshold method in which brightness variation is adapted based on the background is developed. Assuming the brightness variation as a normal distribution, the method estimates the normal range using a single coefficient. By using a single parameter, it is easy to calculate the adaptive threshold for each pixel. Smaller defects at 1/4 the signal level can be detected. Thus, both high sensitivity and easier threshold setup is achieved.

1. Introduction

To achieve and maintain a higher yield, defect inspection of semiconductor wafers is employed during wafer production. Figure 1 illustrates a schematic of a printing circuit. A film material is formed on a silicon wafer. After coating a resist on the material, a mask pattern is exposed and developed. By etching the film through the resist pattern and then removing the resist pattern, one layer of the circuit is made. Semiconductor devices are fabricated by repeating these processes. During fabrication, particles, scratches, and pattern defects sometimes occur for various reasons. Therefore, wafers are usually inspected after each process in the production line. Figure 2 shows scanning electron microscope (SEM) images of defects.

Defects that decrease yield are called critical defects. In efforts to improve yield, it is important to detect critical defects and to devise countermeasures against them [1].

The purposes of wafer inspection are to:
1) Monitor defects stemming from problems with or failure of production equipment;
2) Determine the cleaning cycle of production equipment based on statistical management of wafer inspection results;
3) Predict yield: using circuit design and defect information such as defect number and defect count, Monte Carlo simulation is done to calculate the effect the fabrication process has on yield.

To achieve these purposes in mass production, it is preferable to monitor defects based on defect count than to review each defect, in order to cope with the enormous amount of inspection data. In this case, false alarms can result in misjudgments, so it is necessary to detect minute defects without any false alarms. For example, a technique to control the process by extracting only critical defects using a real-time defect measurement function has been reported [2].

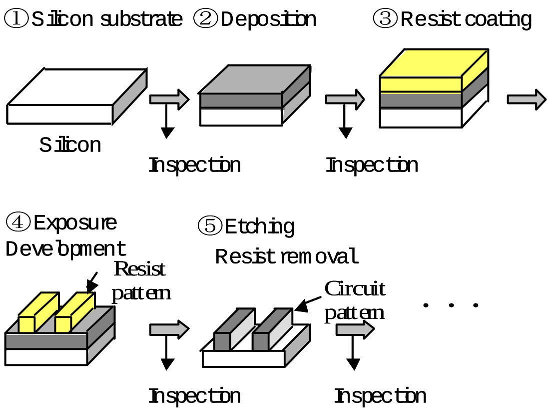


Fig. 1 Semiconductor fabrication Process

High sensitivity inspection technology is necessary all the more as a result of the continuing reductions in size of microfabrication. The capability to detect defects smaller than 100 nm is needed. Additionally, to prevent delays in detecting problems or defects, it is necessary to inspect wafers in many processes, which leads to the need for high speed inspection technology. Moreover, ways to reduce the time and cost of the inspection setup are strongly needed, since the number of parameters for setup are expected to be minimized.

In this article, a method to achieve a statistical threshold that can be set automatically according to the background in dark-field optical inspection is proposed. This method enables high speed inspection. Real-time background noise measurement for each inspection pixel, and appropriate threshold setup using only a single parameter are described.

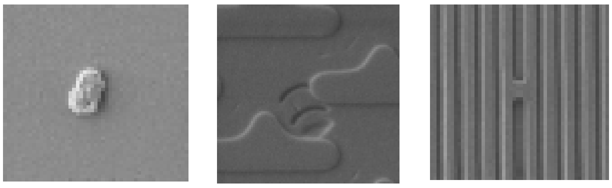


Fig. 2 Defect of Interest (SEM image)

2. Inspection method

The inspection system for semiconductor wafers is comprised of an image detection step that acquires images of the inspection sample, and an image processing step that extracts defects from images.

2.1. Image detection

There are three major image detection methods for wafer inspection: optical dark-field detection, optical bright-field detection, and electron beam detection (Fig. 3).

Dark-field detection, which is targeted in this report, employs oblique illumination and a detector to detect scattered light from the surface of the sample. The detected image is dark because reflected light is not detected, and pattern edges and defects are detected brightly. Because of this optical setup, high sensitivity is easily achieved in inspecting flat wafer surfaces such as wafers after film deposition. Therefore, relatively large pixel size inspection (low magnification inspection), which means high speed inspection, can be performed. Furthermore, for repeated patterns such as memory cells on wafers, spatial filtering techniques are well known [6]. With coherent laser illumination, diffraction at a certain angle is caused by a repeated pattern. Thus, it is easy to block diffracted light using a spatial filter to prevent diffracted light penetrating the sensor. Spatial filtering also enables large pixel inspection. However, scattered light from the pattern edge is detected from a random pattern area, and it is therefore necessary to eliminate the pattern signal (pattern noise) by using image processing.

	Dark-field	Bright-field	SEM
Illumi-nation	Laser	Lamp / Laser	electron
Optics			
Feature	Surface inspection High throughput	High sensitivity Low throughput	Interconnect inspection Low throughput

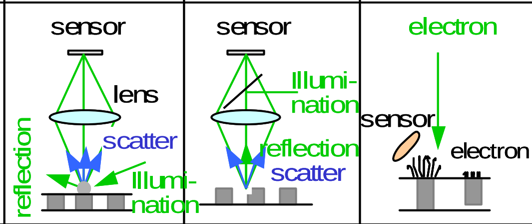


Fig. 3 Image detection methods

2.2. Image processing

The comparison methods below are known to eliminate pattern noise for detecting defects [10][11][12].

(1) Comparison with pattern design data
(2) Comparison with statistical images
(3) Comparison of sample images

In this report, 'comparison of sample images' is adopted. This method is widely adopted for wafer inspection [5][6][7][9]. The same pattern design is formed regularly in each die on the wafer, so die-to-die comparison, depicted in Fig. 4, is commonly used. In die-to-die comparison, defects are extracted after obtaining subtraction image data from neighboring dies. Figure 4 shows a basic configuration, and a series of movements to defect detection is explained as follows. The wafer is placed on a stage that scans at fixed speed. The surface of the wafer is illuminated by a laser at an oblique angle. Scattered light from the surface is collected by an objective lens, which then forms an image at the focal plane. The image is detected by an image sensor and sent to an image processing unit as a sequential digital image. The sequential strip-shaped image with the width of the image sensor is called a swath (Fig. 4). Images between a detected image and a neighboring die image are subtracted to emphasize inconsistent parts. Finally, a threshold is compared with the subtraction image to detect defects. Some inspection tools can measure defect size and/or classify defect category in real time.

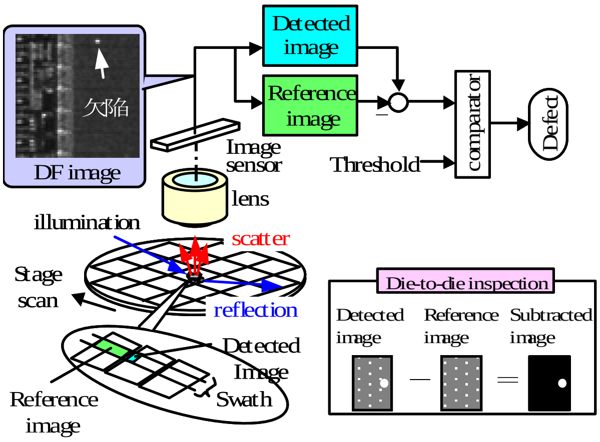


Fig. 4 Die-to-die inspection

3. Threshold setup method

In the following section, a problem in threshold setup and countermeasures for the problem are explained.

3.1. Problem in threshold setup

The scattered light signal of the circuit pattern in a detected image is eliminated to some extent by using die-to-die comparison. However, a defect signals as well as some noise still remain. The noise derived from a sample is usually dispersion of the pattern form or dispersion of a film thickness, for example. Noise coming from the inspection tool is due to misalignment between the reference image and the detected image because of a sampling error and stage vibration. Additionally, speckle noise occurs with coherent laser illumination. Of the above noises, dispersion of the pattern form, dispersion of film thickness, and speckle noise vary in the region inside the die. Figure 5 shows the brightness unevenness that originated in the dispersion of film thickness as an example. Illumination penetrates the transparent film, such as a SiO_2 film. The light scattered by the pattern inside the film reflects on the surface, causing thin film interference; so brightness unevenness occurs when a film thickness is different between neighboring dies. It is necessary to set a threshold value that exceeds the maximum value for every noise so as not to incorrectly detect a normal part as a defect.

Region-based threshold setup is also used [12]. In this method, the region is divided according to the pattern layout or brightness (Fig. 6). The problem with this method is in determining an appropriate threshold for every region. In this report, a method requiring no regional divisions that is adaptive to the dispersion of brightness and can set a threshold easily is proposed. In the proposed method, noise is measured in real time, and a threshold is set appropriately using only a single parameter. The advantage of this method is that not only noise level, but also noise deviation, can be measured using data from multiple dies. The method that employs multiple dies to measure brightness and dispersion in real time is not reported.

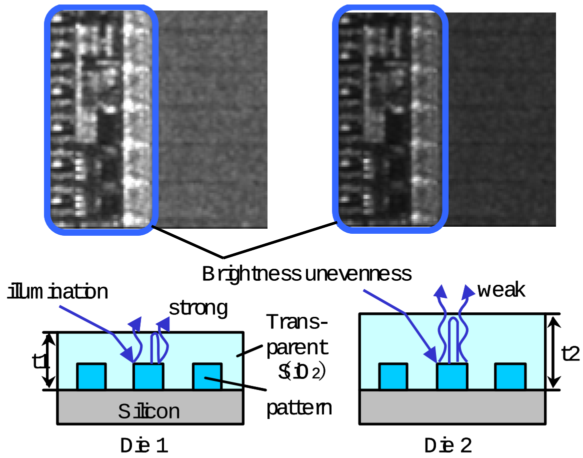
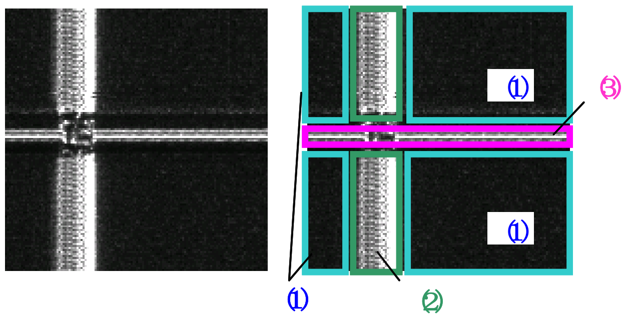


Fig. 5 Brightness unevenness caused by film thickness dispersion



(a) Detected image (b) Region based threshold

Fig. 6 Region based threshold using layout information

3.2. Statistical threshold with background estimation

Figure 7 shows a distribution of brightness skewed corresponding to pixels of multiple dies. Assuming that circuit patterns and inspection tools are ideal, exactly the same images of the die would be acquired. In an actual case, however, there are some differences between these images because of noise. Therefore, the brightness distribution is according to corresponding pixels. As illustrated in Fig. 7, the distribution of brightness differs according to patterns on the wafer. The concept of threshold setting is based on estimating a normal part for each set of skewed pixels in the brightness distribution and then detecting outliers as defects.

There are various methods to define normal range in the brightness distribution; these are shown in Fig. 8. The first method applies a polynomial equation to express distribution. Correlations of polynomials can be calculated by solving a normal equation using the least mean squares method [13]. The normal range is defined as the region where the value of the polynomial equation exceeds a certain parameter. The problem with this method is deciding the appropriate order of the polynomial equation. If the order of the equation is too high, overfitting to the distribution will occur. This means a decrease in generalization ability, which leads to instability. If the order of equation is too low, it will no longer be an approximate expression.

In the second method, a normal distribution is applied to the brightness distribution. Mean and standard deviation are sufficient to express the normal distribution. The normal range of the distribution is defined as the region where the value of the polynomial equation exceeds a certain parameter. The benefits of this approximation are that there are few parameters to set, and the estimation of the tails of the distribution is stable.

The third approach is to estimate the normal range by focusing only on the maximum and the minimum of the distribution and setting an offset to the maximum and the minimum. This approach also has few parameters and is inexpensive to calculate. The problem with this approach, though, is that it is difficult to determine an appropriate offset because the method only focuses on one data value for each end of the distribution.

The fourth method is the percentile method [14]. This method uses cumulative counting. As depicted in Fig. 8, the cumulative area from one end of the distribution is calculated to find the value at which the cumulative area equals a certain proportion of the total area of the distribution; then an offset is set to define the normal range. This method is more stable compared to the third approach, but is much more costly to calculate. Additionally, two parameters—the proportion of the area and the offset—need to be set.

We consider three aspects from a practical use perspective in order to choose a method.

(a) Number of parameters
(b) Calculation resources
(c) Stability of normal range estimation

From these points of view, the normal distribution method is adopted. It is assumed that the variation in brightness will comply with normal distribution as the general error distribution. Population of a mean and standard deviation for each skewed pixel is obtained from multiple dies in the swath.
We named this normal distribution method ‘statistical threshold with background estimation.’
As mentioned in the second section, a threshold is applied to the subtraction image in die-to-die comparison. If a brightness distribution of the skewed pixels complies with normal distribution, the subtraction distribution of skewed pixels for subtraction images complies with normal distribution, too [14]. The threshold calculation step is as follows.

(1) Calculate the subtraction images of the dies in a swath;
(2) Calculate the mean and standard deviations for each distribution of skewed corresponding pixels;
(3) The threshold is:

$$\text{Threshold} = m \pm k \cdot \sigma \quad (1)$$

Coefficient ‘k’ is the single parameter to set the threshold. We call this parameter the ‘threshold coefficient’ in this report. Determining ‘k’ is similar to determining ‘p’ in Fig 8.

Figure 9 shows an example of the distribution of the subtraction images. It is confirmed that the distribution has a single peak and that skewness of the distribution is small. A χ^2-test was done to confirm the goodness of fit to normal distribution.

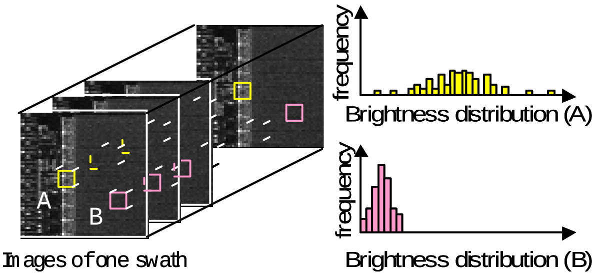


Fig. 7 Brightness distribution of corresponding pixel

3.3. Problem and countermeasure for application

The proposed method estimates the normal range in the distribution. If defect data exists in the distribution, the estimation may fail. To overcome the effect of defect data, the maximum

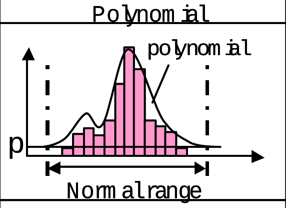
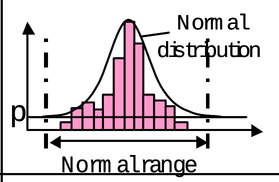
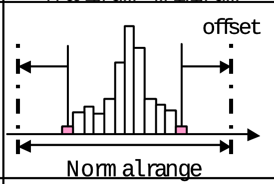
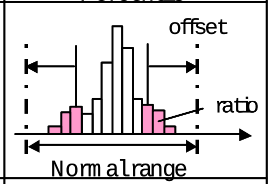
method	Polynomial	Normal distribution	Maximum-minimum	Percentile
Concept image				
Number of parameters	2 (Order of polynomial p)	1 (p)	1 (offset)	2 (ratio offset)
Calculation cost	moderate	moderate	small	large
Estimation stability	unstable	stable	unstable	stable

Fig. 8 Proposed method for threshold decision based on normal range estimation

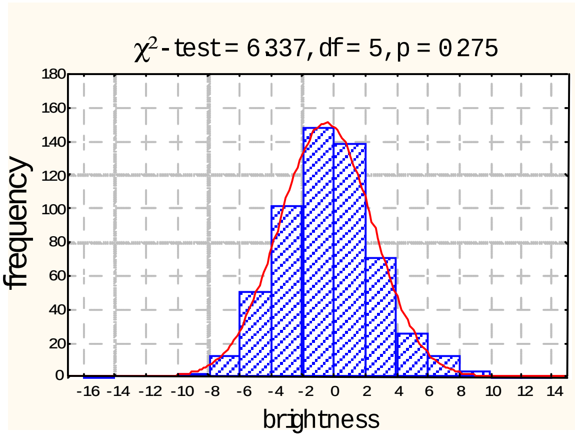


Fig. 9 Brightness distribution of subtracted image

and minimum in each distribution are removed when calculating the mean and the standard deviation. There may be more than one defect datum in the distribution, which will lead to an estimation error. It is assumed that the possibility of the existence of multiple defects in skewed pixels is very low. The reason why both the maximum data and the minimum data are eliminated is that if there are defect signals in a detected image, die-to-die subtraction produces two outliers. The number of data used for calculation is the die count in the swath minus three.

If corresponding pixels have no defect data, eliminating the maximum and minimum data has little effect on calculating the mean and the standard deviation. Also, when there are defect data in the corresponding pixels, the defect data become the maximum and the minimum by die-to-die subtraction. By removing these data, the mean and the standard deviations are calculated correctly.

This method assumes the normal distribution; if the brightness distribution differs vastly, it may cause a false alarm. In this case, the ‘threshold coefficient’ needs to be set larger to improve robustness. By using this method, an appropriate threshold that changes according to the dispersion in brightness of pixels is achieved.

4. Configuration of the experimental apparatus

An experimental apparatus was made to confirm the effectiveness of the proposed method. Image processing was done by software implemented in a personal computer. Figure 10 shows the software configuration. The image processing step consists of a die-to-die comparison stage, and a threshold calculation stage.

In the die-to-die comparison stage, subtraction between dies is calculated sequentially for all images in the swath. In the threshold calculation stage, a threshold for each corresponding pixel is calculated. To increase the number of histogram data, an area of 3 x 3 pixels surrounding the pixel of interest was used. In the threshold calculation, the maximum data (and surrounding 3 x 3 pixels) and the minimum data (and surrounding 3 x 3 pixels) are eliminated to prevent outlier data from affecting the distribution. In the next step, summation, sum of squares, squares of sum, etc., are calculated, and then mean and standard deviations are found. After that, by using predetermined parameter ‘k’, an upper threshold and lower threshold are acquired. Finally, subtracted images of the swath are compared with the threshold to find pixels with values larger than the upper threshold or smaller than the lower threshold; these pixels are defects.

The sequence to calculate a threshold is a simple pipeline procedure without branches or breakpoints. It is easy to implement in a field programmable gate array (FPGA), making it easy to achieve high speed processing.

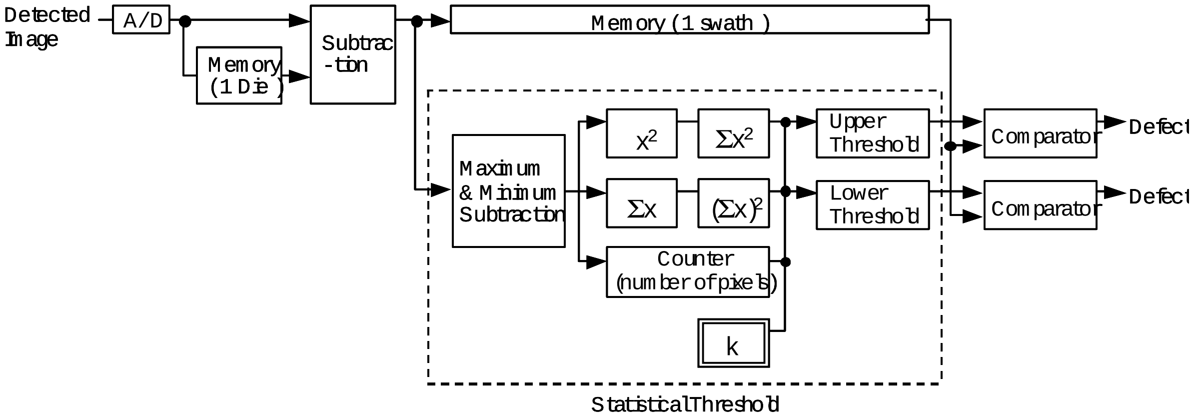


Fig. 10 Architecture of image processing software

5. Procedure to determine threshold coefficient ‘k’

For the inspection of product wafers, it is necessary to set threshold coefficient ‘k’ with the appropriate value. The procedure to determine the threshold coefficient is as follows.

First, an appropriate false alarm rate (the ratio of false alarms to the number of total detected defects) is set in accordance with the purpose of the inspection. A false alarm rate of 20% is acceptable during the ramp up of the production line, whereas only 5% is acceptable during the period of mass production. Next, a wafer to be inspected is prepared, and a trial inspection is done for one swath with an arbitrary value of threshold coefficient k. Subsequently, the coordinate detected as the defect is observed with an optical microscope, or in some cases, with an electron microscope, to confirm the existence of the defect. Then, the false alarm rate for one swath is calculated and compared with an acceptable value to see if it is lower or higher than the acceptable value. If it is lower, the threshold coefficient is decreased; otherwise, the threshold coefficient is increased. After threshold coefficient k is determined for the swath, the whole wafer is inspected, and fine adjustments of the threshold coefficient are done so that the false alarm rate is finally an appropriate value. In this method, k is the only parameter to be determined based on the false alarm rate, so the inspection setup is very simple.

6. Evaluation result

6.1. Experimental conditions

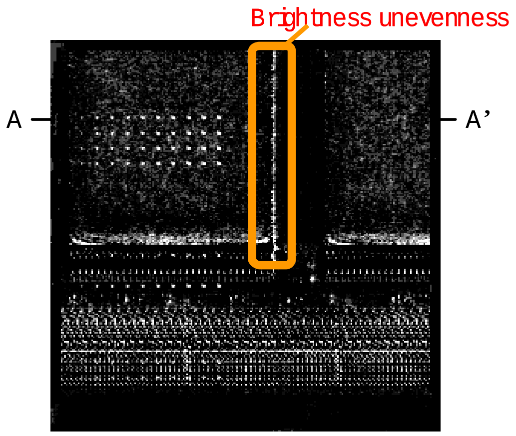
The proposed method was applied to a dark-field detection image of a semiconductor wafer, and the effect was confirmed. The illumination wavelength was 532 nm; detection pixel size was 2.0 µm, the device type was a memory, and the process step was STI-CMP (shallow trench isolation-chemical mechanical polishing) and metal etching.

Generally there are several dozen dies in one swath, but in this experiment, there were ten, and the die size was assumed to be large. In other words, we prepared 10 detection images of a part of the dies, and a threshold was calculated for every pixel. The false alarm rate was set to 0%; in other words, the number of false detections was 0. This is because the evaluation was done using a software program, and the evaluated area was relatively small compared to the whole area of the wafer. Detection performance was evaluated based on the following conditions. An artificial defect that produced changes in the signal level in increasing increments of 10 at each step was made in part of the detected image in a 10 x 10 array, and the defect detection properties were also evaluated.

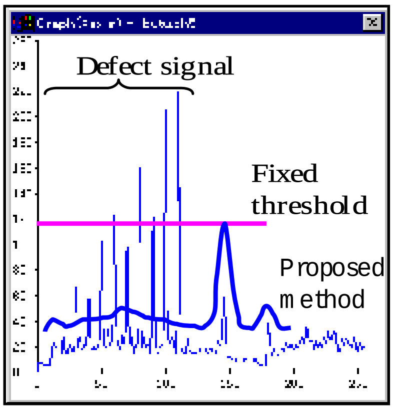
6.2. Experimental result

A sample 256 x 256-pixel subtraction image of the 8-bit gray scale (256 gray scale) and the signal waveform diagram are shown in Fig. 11. The noise of the brightness unevenness region, which is surrounded by a bold line, is high compared with other regions, and because of this, achieving highly sensitive inspection is difficult. This is because a high threshold is needed if the threshold is fixed so as not to detect the brightness unevenness.

Threshold coefficient k was evaluated first. In the STI-CMP process, the value of k that meets a 0% false alarm rate becomes 6. In this case, the defect detection rate is 86%. However, it was found that there were false alarms in the case of k=10 in the metal etching process, as shown in Fig. 12.



Example of evaluated image (subtracted image)

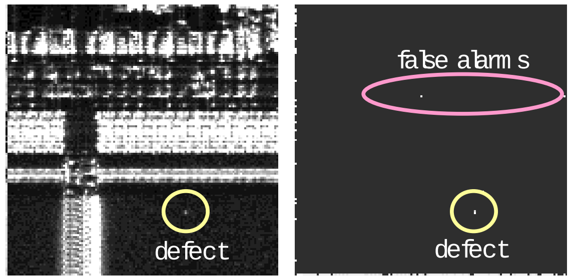


Signal profile (A-A')

Fig. 11 Example of image and effect of proposed method

A lump-shaped change in a form called a grain on a metal wiring pattern occurs in the metal etching process. Grains have no effect on the performance of the semiconductor device; they should be regarded as normal. Grains are detected even if k is enlarged. This is because the frequency of occurrence of grains is so small that it is not reflected on a histogram. To evaluate the effect of the number of data, 5 x 5 pixels surrounding the pixel of interest (250 data) and 7 x 7 pixels surrounding the pixel of interest (490 data) were compared. If the threshold is calculated using the data of 5 x 5 pixels, threshold coefficient k can be lowered with this wafer. When the number of data is small, a histogram of the normal range sometimes follows a t-distribution rather than a normal distribution, so it is important to include an adequate amount of data.

Figure 13 shows the defect detection rate that is improved by using the proposed method. The detection rate for a small signal level is improved in comparison with the former method. A defect at 1/4 the signal level can be detected at the detection rate of 60%. By using the proposed method, high sensitivity is achieved in the region where there is little variation in brightness because a low threshold is automatically applied for that area. Moreover, the threshold can be controlled by the single parameter k, the threshold setup for different devices or different processes. Figure 14 shows an example of a defect detected using the proposed method.



Detected image   Detection result

Fig. 12 Detection result of metal etched wafer

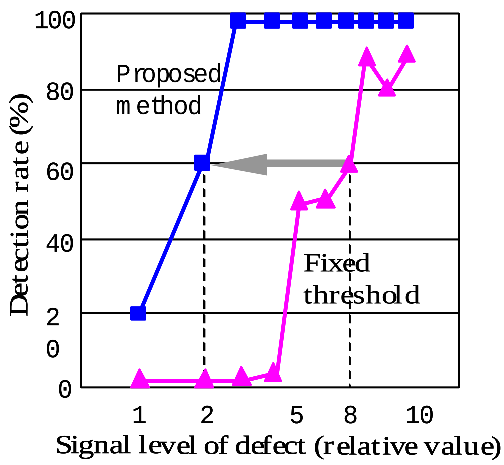
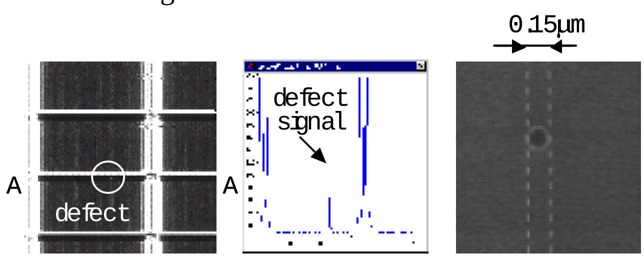


Fig. 13 Defect detection rate



Detected image   Signal profile (A-A')   SEM image

Fig. 14 Detection result
(STI CMP process wafer)

7. Conclusion

A statistical threshold method for semiconductor wafer inspection that can be synchronized to image detection and that calculates the threshold for defect detection by measuring noise dispersion using multiple dies was developed. Dispersion in brightness occurs according to the pattern. It is assumed that the distribution of brightness of corresponding pixels for each die complies with normal distribution, and thus the outlier from the distribution is detected as a defect. It is possible to detect a defect at a smaller signal level (1/4) in comparison with the conventional fixed threshold method. The bottom line of this method is that threshold setup is based on the dispersion of each pixel by using a single parameter; therefore, threshold setup is achieved very easily, which is very important in practical usage. Continuous improvement by using a combination of other proposed techniques is the focus of future studies.

8. References

[1] Smith, J. O. and Abel, J. S., "Bark and ERB Bilinear Transforms," *IEEE Trans. Speech and Audio Proc.*, 7(6): 697-708, 1999.

[2] Kenji WATANABE, Hidetoshi NISHIYAMA, Minori NOGUCHI, Daisuke FUJIKI and Kazunori NEMOTO, "A Proposal of Sensitivity Optimization Method for Dark-field Particle Inspection Technique in Semiconductor Manufacturing," JSPE Trans. Vol.68, No.4: 521-525, 2002 (in Japanese)

[3] http://www.itrs.net/: International Technology Roadmap for Semiconductor, 2005 edition, 2005

[4] Akira HAMAMATSU, Hisae SHIBUYA, Yoshimasa OSHIMA, Shunji MAEDA, Hidetoshi NISHIYAMA, Minori NOGUCHI, "Statistical Threshold Method Based on Background for Semiconductor Inspection," Vision Engineering Workshop Proc., 1-5, 2004 (in Japanese)

[5] Nobuyuki AKIYAMA, Toshihiko NAKATA, Hiroshi MAKIHIRA, Masayoshi SERIZAWA, Hideaki DOI, Fumio MIZUNO, Mari NOZOE, Masami IKOTA, Takuro HOSOE, Shuichi CHIKAMATSU, Takahiro JINGU, Shin ITO and Shigeru ABE, "Development of Particle Detection System using Subtracted Image Signal," JSPE Trans. Vol.58, No.11: 121-126, 1992 (in Japanese)

[6] Nobuyuki AKIYAMA, Fumio MIZUNO, Masami IKOTA and Katumi TAKAMI, "Particle Detection Technology on Wafer Using Spatial Filter," JSPE Trans. Vol.58, No.11: 121-126, 1992 (in Japanese)

[7] Shunji MAEDA, Takashi HIROI and Hitoshi KUBOTA, "Automated Visual Inspection of LSI Wafer Multilayer Patterns Using a Derivative-Polarity Comparison Algorithm," IEICE Trans., Vol.82 D-II, No.1:39-52, 1999 (in Japanese)

[8] Yasuo NAKAGAWA, Workshop for Automation of Visual Inspection, Vol.14, 1999

[9] W.D.Meisburger, A.D.Brodie and A.A.Desai, "Low-voltage electron-optical system for the high-speed inspection of integrated circuits," J. Vac.Sci.Techinol. B10(6), pp.2804-2808(1992)

[10] Yasuo NAKAGAWA and Takanori NINOMIYA, "Visual Inspection Technology for electronic circuit," O plus E, No.132: 138-152, 1990

[11] Julian Powell and Brian Carignan, "CAD Reference for AOI," Printed Circuit Fabrication, Vol. 12, No. 12, pp. 77-93, 1989

[12] J.R.Dralla, J.C.Hoff and A.H.Lee, "Automated submicrometer defect detection during VLSI circuit production," SPIE Proceedings, vol. 775,: 218-225, 1987

[13] For example: David C. Lay, "Linear Algebra and Its Applications (3rd Edition)," 2002

[14] Fro example: John E Walsh, "Handbook of Nonparametric Statistics Investigations of Randomness, Moments, Percentiles, and Distributions," 1962

[15] David Freedman, Robert Pisani, and Roger Purves "Statistics," Third Edition, 1997