Available online at www.sciencedirect.com
SCIENCE DIRECT®
Microelectronic Engineering 77 (2005) 337–346
MICROELECTRONIC ENGINEERING
www.elsevier.com/locate/mee

Defect detection on semiconductor wafer surfaces

N.G. Shankar [a,*], Z.W. Zhong [b]

[a] Department of Engineering, Euro Technology Pte Ltd., 4012, Tech Place 1, Singapore 569628, Singapore
[b] School of MPE, Nanyang Technological University, Singapore

Received 30 August 2004; accepted 14 December 2004
Available online 25 December 2004

Abstract

A template-based vision system for the 100% inspection of wafer die surfaces has been developed. Design goals included a requirement for the detection of flaws as small as two thousandths of an inch on parts up to 8-in. wafer size. Each die is treated as one part of the whole wafer. One of the good dies is trained and kept as template die for the whole wafer. The die physical location data are generated, beforehand, from the wafer map supplied by the wafer manufacturers. A separate software package called the wafer map editor (WME) was developed to generate setup data needed during the runtime of the inspection process. The WME generates an updated wafermap with in-house defects after inspection of the wafer surface. Each unique die pattern in the wafer is defined as an object and these objects are grouped into user-defined categories, such as good die pads and defect die pads. The defect wafer map is used during the runtime inspection of die attach to avoid picking of bad dies.

Keywords: Machine vision; Wafer inspection; Semiconductor; Electronic wafermap; Mean square error

1. Introduction

In the majority of semiconductor assembly, the visual inspection process of the wafer surface depends on manual review by human experts. Since the inspection task requires extreme concentration, the time that an inspector can continue the task is quite limited, and still, it tends to be quite slow and inaccurate. As integrated circuit (IC) feature sizes shrink, semiconductor processes become more complex, and new defect classes become yield limiters. New process technologies like chemical mechanical polishing (CMP) introduce hitherto unknown classes of surface defects [1]. It has been estimated that up to 80% of the yield loss in the production of high-volume, very-large-scale integrated (VLSI) circuits can be attributed to random visual particle and pattern defects [2]. Contamination particles that did not create problems with 1 μm design rules can now be categorized as ‘‘killer

* Corresponding author. Tel.: +65 64559838; fax: +65 64559822.
E-mail address: ngshankar@pacific.net.sg (N.G. Shankar).

0167-9317/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.mee.2004.12.003

defects'' as critical dimensions dip below 0.18 μm, i.e., defects which result in improper electrical device functions [3]. Also the increasing demands for miniaturization, high electrical performance and high I/O pin count has resulted in intensive research and development of advanced IC packages [4,5], which results in much critical inspection task. The continued trend in semiconductor manufacturing towards higher density devices and larger wafer formats is resulting in a greater need for automated yield analysis tools. The increased application of image-based defect detection and review workstations for process monitoring and characterization is generating considerable amounts of data for evaluation by production personnel. This data was necessary to evaluate the state of the manufacturing process and to ultimately improve product yield in a timely manner. Defect yield management tools are beginning to appear in the market, which allow the user to archive and review various permutations of semiconductor defect wafermaps and high-resolution defect image-data [6], but to date their ability to automatically recognize and classify anomalous patterns in the data has been limited. For the past few years, the international technology roadmap for semiconductors (ITRS) [7] has highlighted the increasing need for technologies that address key defect detection and characterization requirements.

The focus of this paper has been on developing an automated method for detecting defects on the wafer die surface using mean square error analysis. Also, the paper focuses to edit the original wafermap and thereby generates final (before die-attach) wafermap of each and every wafer under inspection. Fig. 1 shows a scenario for collecting and analyzing defect data in a production environment. As wafers exit a fabrication process (e.g., fabrication process A), wafermap data were generated by an in-line defect detection workstation generally incorporating a microscopy or light scattering system. Once the instrument has scanned a wafer, its electronic wafermap is moved to a yield management system for prediction or to alarm operators when process is going out of specification or when maintenance or calibration must be scheduled.

The information in the wafermap consists of detected defect coordinates as well as process information such as step, layer, and product. An operator viewing a plot of the coordinate points during analysis typically observes the existence of patterns in the wafermap data manually. Automation tools for wafermap analysis currently use simple nearest-neighbor clustering [8] techniques to group data primarily into ''random'' or ''grouped'' events, but specific classifications are not obtained.

2. The inspection task

Defect detection in semiconductor assembly manufacturing has become increasingly important over the past several years as a means of quickly understanding and controlling contamination sources and process faults, which impact the product yield. To address the above issue, automation technologies in defect detection and review are being developed by universities, laboratories, industry, and semiconductor equipment suppliers. Several techniques were adopted, such as automatic defect classification (ADC) on the sub-die or defect level and spatial signature analysis (SSA) on the whole-wafer level [9]. One of the few commercially available ADC systems extracts numerous spatial and textural features from the defect region, and applies a set of user defined

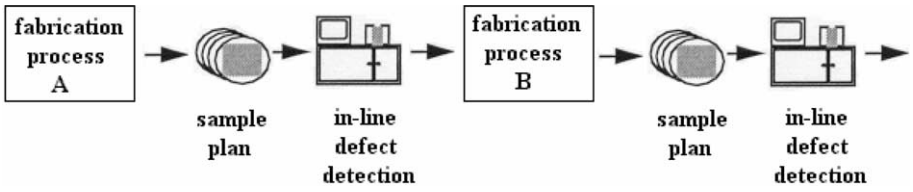


Fig. 1. Schematic representation of a scenario for collecting and analyzing defect data in a production environment.

fuzzy predicates for defining the defect classes. The main draw back of ADC is the time consumption and also there is no reference pattern for the whole process, i.e., it considers neighboring die as reference for the die under inspection. On the other hand, SSA has no relation with the defect data generated by local anomalies caused during wafer mounting, dicing, imbedded particle contamination, etc.

Defect detection is performed by directly comparing the two complex wavefronts taken from corresponding fields of view from adjacent die on the wafer. Difference images can be computed as either amplitude or phase or as a composite difference. In order to compare two images from different die, the pixel values must be aligned in the x- and y-directions (spatial registration) and matched in terms of their overall intensity and phase offset (normalization). Spatial registration of the complex images is performed via an automatic registration algorithm described previously [10]. There are some semiconductor inspection applications, however, in which a reference image is either unavailable or is of such poor quality that it cannot be used as a reliable basis for comparison. In the event that the defects need to be redetected at a future date, the lack of a reference image dictates that a non-referential defect detection (NRDD) approach [11] can be employed. For example, if a yield engineer sees a new defect being detected by the inspection system, his first goal will be to determine the source of the defect. One way to determine the source would be to search the historical defect image database to find previously diagnosed cases in which the defect resembles the new one. Before this database can be searched for defects with similar appearance using a technique, such as content-based image retrieval, [12], the defects within the database must first be redetected. Finding defects without the aid of a reference image is relatively easy for a human, but is an extremely challenging task for a computer vision system. Automated techniques that have been developed to date are highly application-specific in that the image processing techniques used are tuned for the specific object and background that are expected to be present in the image [13]. Several promising well-known techniques, such as digital holography [10], digital shearography [14], and semiconductor neural networks system have been developed and reported by several research groups [15] for classification of the feature vectors and the defect classes. Rule-based inspections of semiconductor wafer surface have been reported [16,17].

Basic studies reveal that each defect object has different surface intensity, shape, color, texture, etc. Each and every time the user has to teach the system to identify and recognize the defect objects by some means. Second, the user has to teach the system for various matrix tolerances and it should be able to tell the user about the shape, size, and type of defects and how to segregate into various classes for future references. As manufacturing processes mature, the occurrence of anomalies arises from particle contamination, mechanical damage and process variations due to improper equipment calibration or miss-out calibration or poor maintenance of the equipment. Equipment with implemented defect pattern recognition algorithms had no prior knowledge of newly developed defect events and also the accuracy of detecting equipment slowly degrades due to aging. Since, most of the front-of-line process equipment in the semiconductor assembly accustomed to the defect data supplied by the wafer manufacturer as wafermaps. These wafermaps have no prior knowledge of the defects generated during assembly process. Hence, there is a necessity to detect these defects along with wafer manufacturer defect data (wafermap) before the die-bond process in the semiconductor assembly. At last, the speed of processing the defective pattern utilizes much of the system operating time.

The increased number of defect types drives up the cost of fault reduction. Since yield engineers have observed that different types of defect inspection equipment capture some defect types more effectively than they do others, companies are forced to buy many types of defect inspection tools in order to identify most potential sources of electrical faults. As process complexity increases, defect detection may pose a daunting economic challenge to all, but the largest semiconductor manufacturers and up to date no body have found a final solution to in-line inspection. In general, it

is difficult to detect defects because of the inherent variability of wafers in the assembly line. Although much of the work reported above has yielded results that are good in a quantitative sense, evaluation suggests that the defect analysis results could be improved in a qualitative sense by refining the resulting shapes and extent of detected defect regions in the images.

For example, small spurious defect regions may have only a small effect on pixel-wise statistical classification accuracy, but are objectionable to the human observer. In many cases, such regions can be identified and eliminated easily. Fig. 2 shows representative examples of the variety of defect images in semiconductor manufacturing. Figs. 2(a)–(d) show examples of individual defect pattern identified using optical and electron microscope by manual inspectors in the assembly area. Hence, it is clear that the defects that arise after wafer die sawing is discarded during die bonding. The goal of this project is to provide an on-line system that can perform 100% inspection of up to 8-in. wafer. The system is required to catch flaws that are unacceptable in the semiconductor industry, and yet allow non-critical variations to pass.

2.1. Implementation

The system consists of a frame grabber with 768×576 PAL/SECAM resolutions and a monochrome camera with a magnification of 5×. The runtime inspection software is written in Visual Basic Language. The runtime system is controlled by an industrial PC running an operator interface written in Visual Basic. A separate software package called the wafer map editor (WME) was developed to generate setup data needed during the runtime inspection process. The wafer map editor includes the location of new defects found during inspection, to the original wafermap supplied by the wafer manufacturer.

2.2. Camera resolution

The task was to determine the minimal camera resolution that would be required. Given the size of the die pattern and the relatively tiny size of the potential defects, cost would rise prohibitively if unnecessary resolution was employed. Through a series of experiments and demonstrations, it was determined that a 2-mil per pixel resolution could do the job. Below the Nyquist sampling limit, use of this camera resolution leads to the detection of 2-mil high-contrast defects most of the time [18]. The defect signal rapidly approaches the noise level as the size drops below 2 mil. The possibility of using a line scan camera together with a precision scanning stage was investigated, but we learned that vibration and scanning stage durability became over riding issues. We decided to use a 768×576 resolution camera mounted on a linear X- and Y-axis stage to acquire images of wafer under inspection.

Illumination was provided by a pair of fluorescent lamps located on each side of the part being inspected. High-frequency ballasts are used to power up the lamps. Moveable shades are added to each lamp fixtures to flatten the illumination field. The system sensitivity must be changed for each part type to provide maximum contrast. Camera registration was greatly improved if the image array is

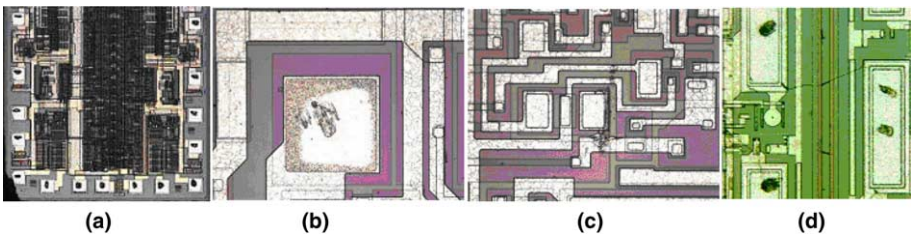


Fig. 2. Examples of various defect image patterns identified by manual inspectors in the assembly area, (a) chipping defect, (b) metallization peel-off defect, (c) scratch defect, and (d) bridging defect.

adjusted to be as nearly parallel to the plane of the part as possible. A white sheet with an 11-by-15 array of filled circles at 1-in. center-to-center spacing is placed on the inspection area.

2.3. The mask image

The mask image in Fig. 3 is derived by edge detection of the reference die (template) beforehand. The absolute difference image is obtained by subtracting the image of the test die from the template and the difference image is multiplied by the mask image to reduce unavoidable pixel differences that arise from variety of sources. When the mask image is black, the non-zero pixel differences are reduced to zero.

The edges of the die are extracted using the canny operator [19]. Edge differences arise from slight misregistrations that exist between the test image and the template. The mask is also used to eliminate acceptable defects in the template, since a perfect template in reality does not exist. Finally, masking is added to eliminate differences due to other marks that are present as part of the manufacturing process.

3. Referential inspection

The use of the "golden" template reference-based inspection was an obvious choice for this inspection. Images of die parts are precisely aligned with a flawless template and are compared by pixel subtraction. Two images that are identical will yield an image containing all zero pixels. However, any differences between the template and the test image will appear as non-zero pixels. The entire process is summarized in Fig. 4.

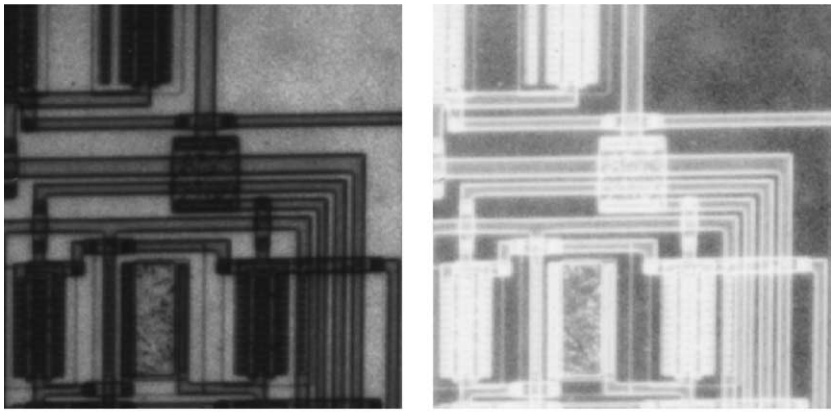


Fig. 3. Shows the extraction of mask image from the reference die pattern.

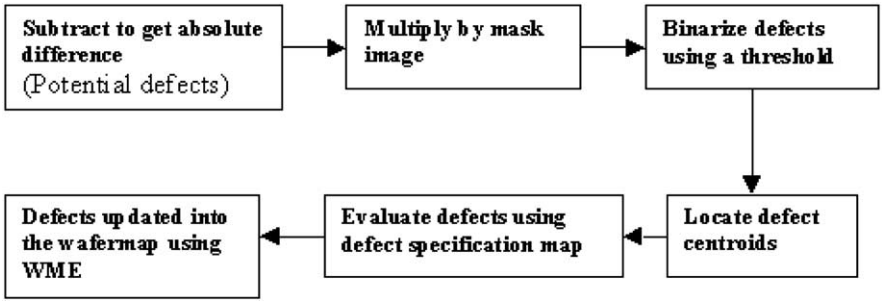


Fig. 4. Summary of the entire inspection process.

3.0.1. Experiment

Let $Y = \mathrm{dct2}(X)$ returns the discrete cosine transform of X, where dct2 is the 2D version of discrete cosine transform. The matrix Y is the same size as X and contains the discrete cosine transform coefficients. $Y = \mathrm{dct2}(X,[M,N])$ or $Y = \mathrm{dct2}(X,M,N)$ quantizes matrix X with zeros to size M-by-N before transforming. If M or N is smaller than the corresponding dimension of X, dct2 truncates X. Reference circuit image is matched to the test image based on the mean square error analysis within the specified range (industry standard $1/10^6$). The algorithm runs until the correction between iterations is below a limit value or the mean squared error does not change sufficiently. A maximum number of iterations terminate the algorithm if convergence is not reached. The result is valid if the mean square error is less than the projected error value (10^{-6}). Otherwise, the result is invalid and the particular test die is considered a defect die whose position is updated into the wafermap with the help of wafer map editor.

3.0.2. Sum of squares due to error

This statistic measures the total deviation of the test image values from the reference image re-

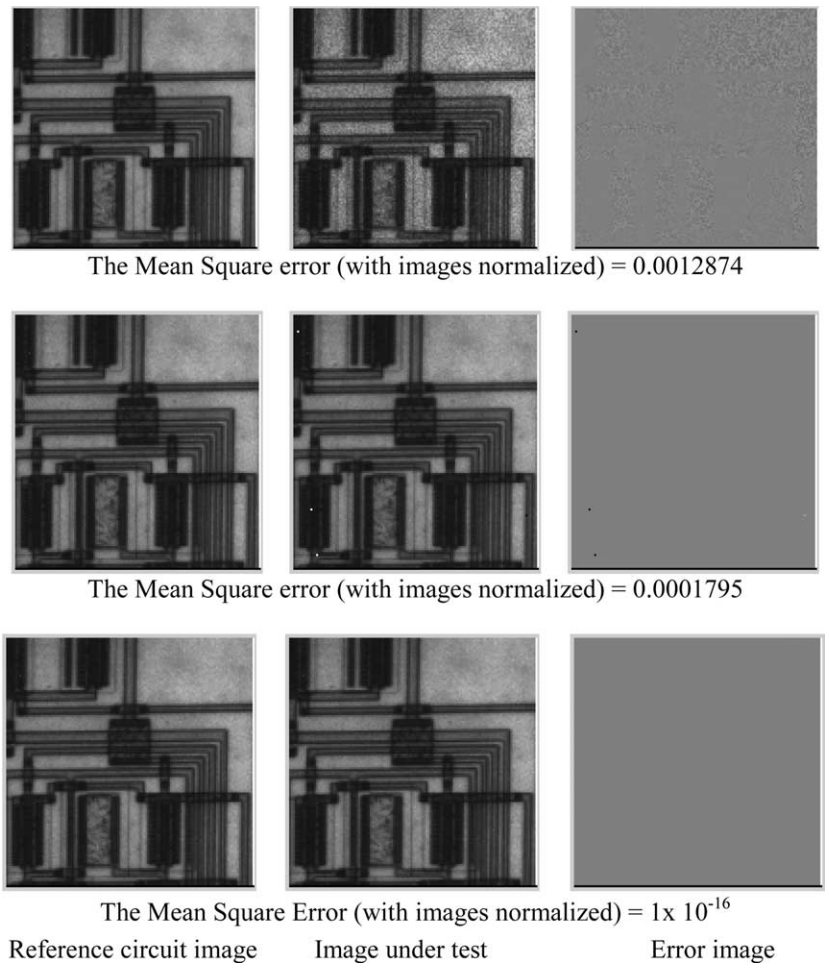


Fig. 5. The visual image slides that illustrate the different error images extracted using mean square error principle.

sponse values. It is also called the summed square of residuals and is usually labeled as SSE and it is expressed according to [20] as

$$\mathrm{SSE} = \sum_{i=1}^{n} w_i (y_i - \bar{y}_i)^2, \quad (1)$$

where n is the number of data points included in the image. The error for the ith data point is defined as the difference between the test image response value y_i and the reference image response value $\bar{y}_i$. Then the mean square error is given as

$$\mathrm{MSE} = \frac{\mathrm{SSE}}{p}, \quad (2)$$

where p is the number of independent pieces of information involving the n data points that are required to calculate the sum of squares. Usually p is the product of the size of reference image.

3.1. Rule-based defect specification

In addition to masking, the system tolerates non-critical flaws by using rule-based defect specification. Differences between the test image and the unmasked template are evaluated with respect to position by using defect map. In the setup phase, the WME is used to assign a number to each object or die in the wafer. The defect map overlays the template so that every pixel of the object or die in the defect map is assigned the same object identifier. When a significant difference "blob" is detected, its centroid is located in the defect map so that its object identifier may be extracted.

The identifier then serves as a key to lookup the defect specification rule for the wafer die image. For example, if the defect were located in a bond pad of the die surface, the rule would probably indicate that this defect is significant and should be called to the operator's attention. The rule would simply state that one or more defects of any size found in a bond pad of this type should be considered to be significant. On the other hand, a small void located in a die pad region would be registered for that particular die pad but would otherwise be ignored. For the sake of this example, the rule would state that up to three voids are allowed as long as the accumulated area does not exceed 10% of the die pad area.

The system then calculates the position and orientation information from an image of the dot pattern acquired by the camera. In particular, the distance of the camera from the sheet and the vertical off-axis angle are used to ensure that the camera is aligned within tolerance for the particular image under consideration [21].

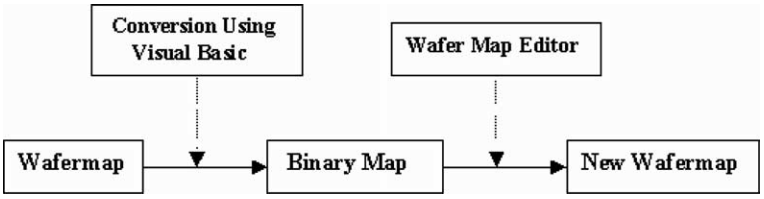
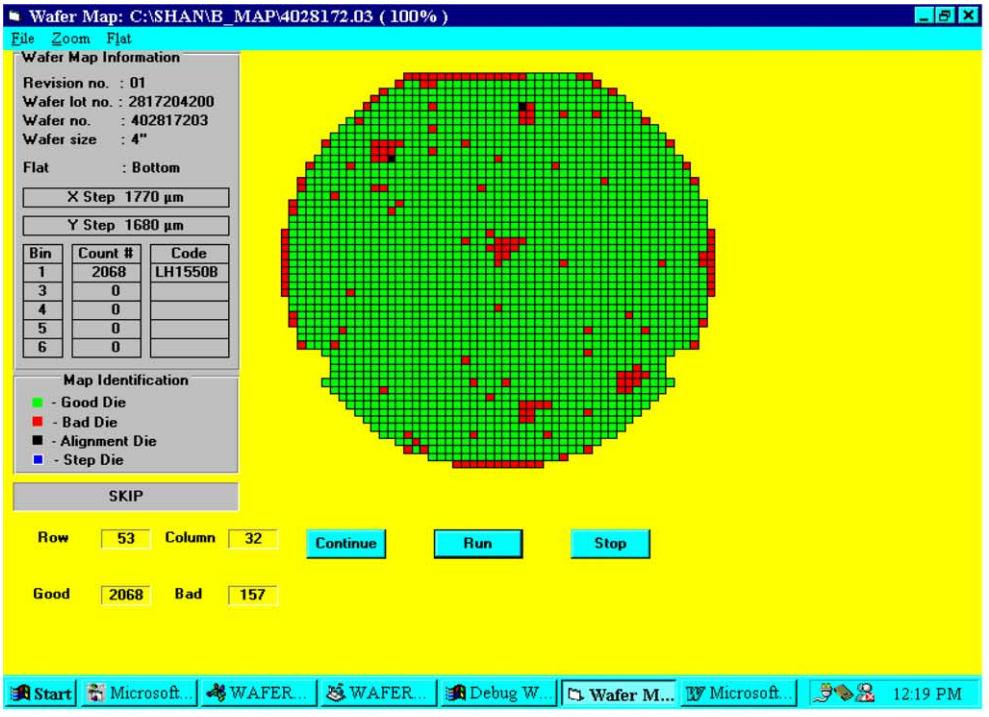


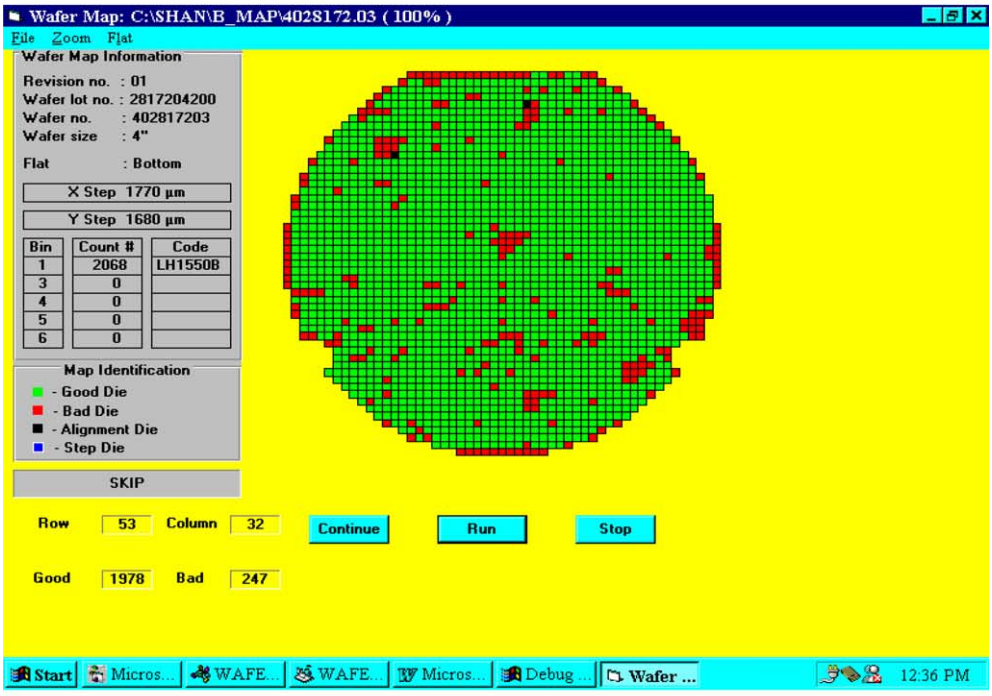
Fig. 6. The wafer map editor provides the runtime inspection to generate new wafermap with part specific data.

Table 1
Testing parameters and results of application of the in-line inspection to a variety of data sets

Wafer size	100 mm	125 mm	150 mm	200 mm
Die size (square)	0.2–18 mm	0.2–18 mm	0.2–22 mm	0.2–25 mm
Street Width (1 mil = 25.4 μm)	2 mil	2 mil	3 mil	3 mil
Wafer thickness	7–20 mil	7–20 mil	7–20 mil	7–23 mil
Inspection time per wafer	10–17 min	10–17 min	12–18 min	12–20 min
Die size tested (square)	1.2–12 mm	1.2–12 mm	1.6–12 mm	2–20 mm
Targeted defects	Foreign particles, saw into metal, crack, scratches, chip die, glassivation void, diffusion fault and discolored bond pad.			



(a) Wafer surface before inspection



(b) Wafer surface after inspection and editing

Fig. 7. (a) Wafer surface before inspection. (b) Wafer surface after inspection and editing.

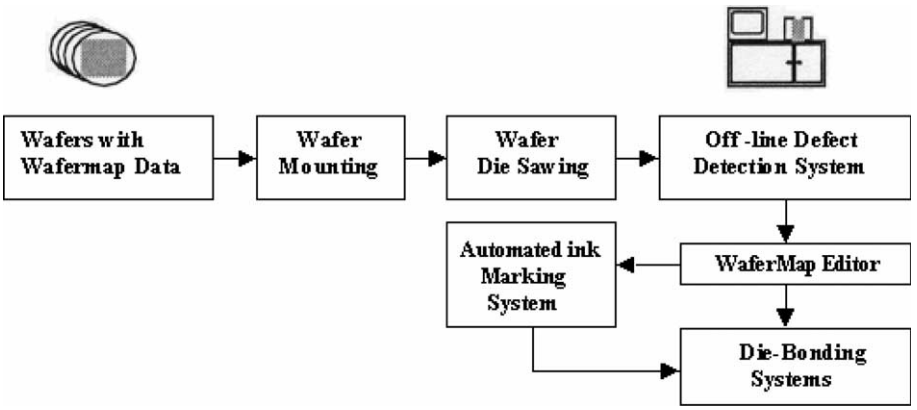


Fig. 8. Typical application of the proposed defect detection system.

The measurement method uses a 2D discrete cosine transform (DCT) mean square analysis. The reference circuit image (template) and the test image are transformed in 8-by-8 blocks. The visual image slides that illustrate the different error images extracted using mean square analysis are shown in Fig. 5. The error image, i.e., the difference between the reference circuit image and the test image is also displayed.

4. The wafer map editor

The WME provides the runtime editing of the original wafermap with most of the required preprocessing data. These data are derived from the wafermap supplied along with the wafers. The information generated by the WME is summarized in Fig. 6.

Processing of the wafermap to a binary map is performed using visual basic software and from the binary map it is known that the good die pattern is marked as binary one (1) and the bad die pattern is marked as binary zero (0). Once, a die pattern is considered as defect using mean square error then the particular die is marked as binary zero using wafer map editor dialog screen. Table 1 summarizes the testing parameters and results obtained from in-line inspection to a variety of data sets, including the die size, inspection time, targeted defect types, and wafer thickness. Fig. 7(a) shows the wafer surface before inspection and Fig. 7(b) shows the final wafermap of a typical 4-in. wafer after inspection and editing using wafer map editor dialog screen. By utilizing the proposed detection system, one can replace the unnecessary verification of defects in die-bond machines thereby improving the assembly line efficiency. Fig. 8 shows a typical application of the proposed system.

5. Conclusions

In this paper, we have described a vision system in terms of its hardware modules, as well as the image processing algorithms that it utilizes to scan the images and to detect defects from the wafer surface. In addition to being tested in a laboratory environment, a prototype of this system was constructed and deployed to a die attach system, where its performance was evaluated under realistic conditions to detect defects after wafer dicing, and it was found that the vision system was able to successfully detect non-uniformities. The experimental results have proven to be good, statistically for the projected mean square error for several species of wafer die sizes.

In summary, a template-based vision system for the 100% inspection of wafer die surfaces has been developed. Design goals included a requirement for the detection of flaws as small as two thousandths of an inch on parts up to 8-in. wafer size.

The system is defect tolerant through the use of defect specification rules that can help the system distinguish between critical and non-critical regions on the wafer die region of interest.

References

[1] C. Dennison, MICRO 16 (2) (1998) 31–41.
[2] H. Miyashita, et al., J. Vac. Sci. Technol. A 17 (3) (1999) 1066–1070.
[3] L. Milor, G. Hill, Y. Peng, in: Proceedings of the International Symposium on Semiconductor Manufacturing, 1999, pp. 99–102.
[4] Z.W. Zhong, K.S. Goh, J. Electron. Manuf. 10 (2) (2000) 89–96.
[5] Z.W. Zhong, K.S. Goh, J. Electron. Manuf. 10 (4) (2000) 211–217.
[6] K.W. Tobin, et al., J. Vac. Sci. Technol. A 17 (4) (1999) 1369–1376.
[7] International Technology Roadmap for Semiconductors, 2000 Update, Semiconductor Industry Association. Available from: <http://public.itrs.net/>.
[8] T.P. Karnowski, S.S. Gleason, K.W. Tobin, SME Technical Paper MS00-240, Society of Manufacturing Engineers, Dearborn, MI, 2000.
[9] K.W. Tobin, F. Lakhani, T.P. Karnowski, Proc. SPIE 4692 (2002) 46–53.
[10] X.L. Dai, M.A. Hunt, M.A. Schulze, in: Machine Vision Applications in Industrial Inspection XI, Proceedings of SPIE, vol. 5011, Santa Clara, CA, 23–24 January 2003.
[11] S.S. Gleason, R.K. Ferrell, T.P. Karnowski, K.W. Tobin, Proc. SPIE 4692 (2002) 61–71.
[12] T.P. Karnowski, K.W. Tobin, L.F. Arrowood, R.K. Ferrell, J.S. Goddard, F. Lakhani, in: SPIE's 13th International Symposium on Electronic Imaging: Science and Technology, San Jose Convention Center, January 2001.
[13] P. Xie, S. Guan, Mach. Vision Appl. 12 (3) (2000) 149–156.
[14] Ganesha Udupa, B.K.A. Ngoi, H.C. Freddy Goh, M.N. Yusoff, Meas. Sci. Technol. 15 (2004) 35–43.
[15] K. Kameyama, Y. Kosugi, T. Okahashi, M. Izumita, IEICE T. Inf. Syst. E81-D (11) (1998) 1261–1271.
[16] N.G. Shankar, Z.W. Zhong, in: The Fourth International Conference on Control and Automation (ICCA'03), Montreal, 2003, pp. 752–755.
[17] N.G. Shankar, Z.W. Zhong, in: ICMAT, Symposium L, Singapore, December 2003.
[18] IEEE Broadcast Technology Society, IEEE Std 208-1995, IEEE Standard on Video Techniques: Measurement of Resolution of Camera Systems, 1993 Techniques, IEEE, 1995.
[19] Image Processing Toolbox, MATLAB, Version 3.2, Release 13, June 28, 2002.
[20] N.R. Draper, H. Smith, Applied Regression Analysis, third ed., Wiley, New York, 1998.
[21] International Standards Organization. ISO 12233: 2000, Photography – Electronic Still-picture Cameras – Resolution Measurements, ISO, 2000.