

Detection of semiconductor defects using a novel fractal encoding algorithm

Shaun S. Gleason^{*a}, Regina K. Ferrell^a, Thomas P. Karnowski^a, Kenneth W. Tobin^a

^aOak Ridge National Laboratory^{**},

ABSTRACT

This paper introduces a new non-referential defect detection (NRDD) algorithm for application to digital images of semiconductor wafers acquired during the manufacturing process. This new algorithm is composed of two major steps: (1) defect detection via the use of fractal image encoding and (2) enhanced defect boundary delineation using active contours. One primary application for this technology is the redetection of defects within archived databases of historical defect imagery. Defect images are commonly stored by semiconductor manufacturers for future diagnostic purposes, but reference (golden) images are usually unavailable. The ability to automatically redetect a defect is crucial in an automated diagnostic system that uses the historical defect images for defect sourcing. Results are presented for four large data sets of semiconductor images. Three of these data sets are composed of scanning-electron microscope (SEM) images and the fourth contains optical microscope images. Performance criteria were created that score the NRDD segmentation result as a percentage based on a comparison to a manually outlined version of the defect. The overall NRDD score across all four databases ranged from 50% to 84% using the same set of manually-determined parameters on all images within each database. By using an automated parameter setting algorithm these performance values improved to 57% to 92%. The NRDD algorithm performance depends, in part, on the size of the defect and the level of complexity of the background of the semiconductor image.

Keywords: Semiconductor defect detection, fractal encoding, active contours, snakes, data reduction, computer vision

1. INTRODUCTION

Computer vision technology has enabled tremendous improvements to be made in the field of semiconductor manufacturing. Perhaps more than any other industry, semiconductor manufacturing has benefitted from automated inspection technologies which are used to detect defects (contamination, lithography errors, etc.) during the fabrication process. The most mature and effective algorithms employed in computer vision systems for defect detection are reference-based in that they rely on the availability of a reference image (or golden image) to use as a basis for comparison. Since a semiconductor wafer typically contains many copies of the same electrical component (each component referred to as a “die”) laid out in a matrix pattern, a reference image for one die is often obtained by acquiring an image of the neighboring die. Differences between the current semiconductor image and its corresponding reference are labelled as defective areas. In an oversimplification of the process, these differences are identified by aligning the two images and then subtracting them from one another to create a difference image. This difference image is typically filtered and then thresholded to reveal the defective regions on the semiconductor wafer¹.

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. For example, many semiconductor manufacturers store large data bases of historical defect imagery to be used for diagnostic purposes. For a variety of reasons including excessive storage requirements, reference images are not always saved with these defect images. 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 must 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

*Correspondence: Email: gleasonss@ornl.gov; Telephone: 865 574 8259; Fax: 865 576 8380; <http://www-ismv.ic.ornl.gov/~gleason>; Oak Ridge National Laboratory, PO Box 2008, Bldg. 3500, MS 6010, Oak Ridge, TN, USA 37831.

**Managed by UT-Battelle for the U.S. Department of Energy.

with similar appearance using a technique such as content-based image retrieval², the defects within the database must first be redetected. Manual delineation of the defects is one possible solution, but is ineffective when large quantities of defects must be detected or if the detection is needed within an otherwise automated inspection system. 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 image³.

This paper introduces an NRDD algorithm that leverages the concept of image self-similarity. Self-similarity is an image quality that is exploited by fractal image compression schemes in that a subregion in one area of an image is described by a mathematical transformation of another subregion within that same image. Storing this transformation requires less space than storing an image subregion, therefore data storage requirements are less and image compression is achieved. In this application, the goal is not image compression—it is defect detection—but the approach has commonalities with image compression. The process of breaking the image up into different subregions and then determining the set of affine transformations that best map one region to another is called *fractal encoding*. The idea behind this NRDD algorithm is that a defect will appear different than the normal image background structure, and, hence, it will be difficult to find an affine transformation that effectively maps another image subregion to the current defective one. A similar approach has been previously developed for computer-aided diagnosis of mammographic images^{4,5}, but a few significant changes in the algorithm were implemented specifically for application to semiconductor images. An image subregion containing a defect is flagged for post-processing as a defective subregion. We call this flagged subregion a *focus-of-attention region* (FAR). The post-processing employed here is the application of active contours, or snakes, to the defect image subregions to more finely delineate the defect boundary.

Sections 2 and 3 introduce fractal encoding and active contours, respectively, which are two of the steps within the NRDD algorithm. Section 4 present the results of the application of NRDD to a large set of defect images, and the last section summarizes the effectiveness of NRDD and discusses its applicability to semiconductor defect detection problems.

2. FRACTAL ENCODING AND FAR GENERATION

Fractal encoding can be used to flag all structures that appear to be different from the normal background structures within an image. As mentioned in the previous section, fractal image encoding is the first of two steps executed in performing fractal image compression. Here, we describe fractal encoding only in terms of its relationship to FAR generation. The theory of the encoding process is only briefly described here, so the interested reader is referred to prior publications^{7,8} for a more detailed treatment of the subject matter. Fractal compression and fractal encoding exploit the property of self-similarity of fractal objects. Exact self-similarity means that the fractal object is composed of scaled-down copies of itself that are translated, stretched, and rotated according to a prescribed mapping or transformation (Fig. 1). If the map is contractive (i.e., it always brings points closer together), then the Fixed-Point Theorem guarantees that the generated fractal object (attractor) is unique and independent of the choice of the initial object. Therefore, given that one is dealing with a simple transformation, compression is achieved when the coefficients of the map are stored in place of the attractor.

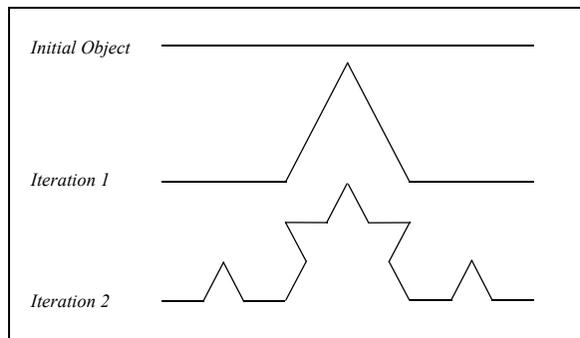


Figure 1. Iterative generation of a self-similar fractal curve. In each iteration, the mapping consists of scaling by a factor of 1/3, rotation by plus or minus 45 degrees, and translation.

An example of a commonly employed mapping is the affine transformation. This mapping is capable of generating a variety of attractors by translating, scaling, rotating, and adjusting the contrast and brightness of an input image. In general, real images do not exhibit exact self-similarity, but they do exhibit *partitioned* self-similarity. That is to say that instead of being formed of copies of its *whole* self, the image is composed of properly transformed *parts* of itself. Fractal encoding exploits the property of “partitioned” self-similarity of images. This means that instead of being formed of copies of its whole self (as in exact self-similarity), the image, in an approximate sense, is composed of transformed parts of itself. In computing the coefficients of this transformation or map, it is assumed that because of the notion of partitioned self-similarity, each subregion of the image can be described (in the sense of minimizing a dissimilarity metric) in terms of another. The former subregion belongs to the range pool, R , while the latter belongs to the domain pool, D , of the map. If a given subregion in D cannot be mapped to any region in R (i.e., their measure of dissimilarity is above a specified threshold, T), then R is further partitioned into smaller subregions. This process continues recursively until either a similar subregion from D is found or a specified maximum level of partitioning, L_{max} , is reached.⁶ The range pool partitioning scheme can take on a variety of forms. For this research we have employed the commonly utilized quadtree partitioning scheme.⁸

In this study, we showed that during the fractal encoding process, for subregions in R that contain semiconductor defects, L_{max} will be reached. The reasoning is that the visual appearance of abnormalities (e.g. particles) is different from that of the coexisting normal structures. Therefore, subregions in D that can be mapped to those areas in the image with defective regions are expected to be nonexistent. A simple example of a semiconductor image that helps to illustrate this process is shown in Fig. 2. Subregions for which L_{max} is reached along with their 8-neighbors make up the FARs; see Fig. 2(c), (d). Four examples of FARs generated for more complex semiconductor images are shown in Figs. 3 and 4.

3. ACTIVE CONTOURS

The fractal encoding and quadtree partition selection process identifies candidate FARs that contain defective image regions. As can be seen in Fig. 2(d), the quadtree partitioning scheme generates a FAR that consists of a collection of adjacent squares that cover the defect. The precise location of the defect boundary is still unknown, however. Active contours, or snakes, are used to automatically delineate the boundary of the defect.

The original research on active contours by Kass, Witkin, and Terzopoulos⁹ is considered by most to be the foundational work in the field. Active contours are constrained by local smoothness parameters and have no overall shape constraints. These researchers coined the internal and external energy concepts that are still widely used in current algorithm development strategies for many boundary-based segmentation techniques. Their energy function is modeled by internal and external energy forces as follows:

$$E_{snake} = \int_0^1 E_{snake}(v(s)) ds = \int_0^1 (E_{int}(v(s)) + E_{ext}(v(s))) ds \quad 1$$

where

$$E_{ext}(v(s)) = E_{image}(v(s)) + E_{con}(v(s)), \quad 2$$

$v(s)$ is the parametrized contour $[x(s), y(s)]$, E_{int} represents the internal energy of the spline due to bending, E_{image} represents forces imposed by the target image, and E_{con} represents external constraint forces that the user may impose. For snakes, the internal energy is related to the smoothness of the active contour, and the external energy is based on edge information in the target image. An unique energy formulation was more recently developed by Xu and Prince¹⁰ and that is the one used in this work.

Snakes were designed to be interactive and require good initial guesses to converge to the correct solution, because they tend to become trapped in local minima. This is especially true when the image is cluttered with spurious edges. To effectively initialize the snake position for this defect detection application, we utilize spatial information derived from the FAR. An example result of applying the snake algorithm to the image from Fig. 2 is shown in Fig. 5.

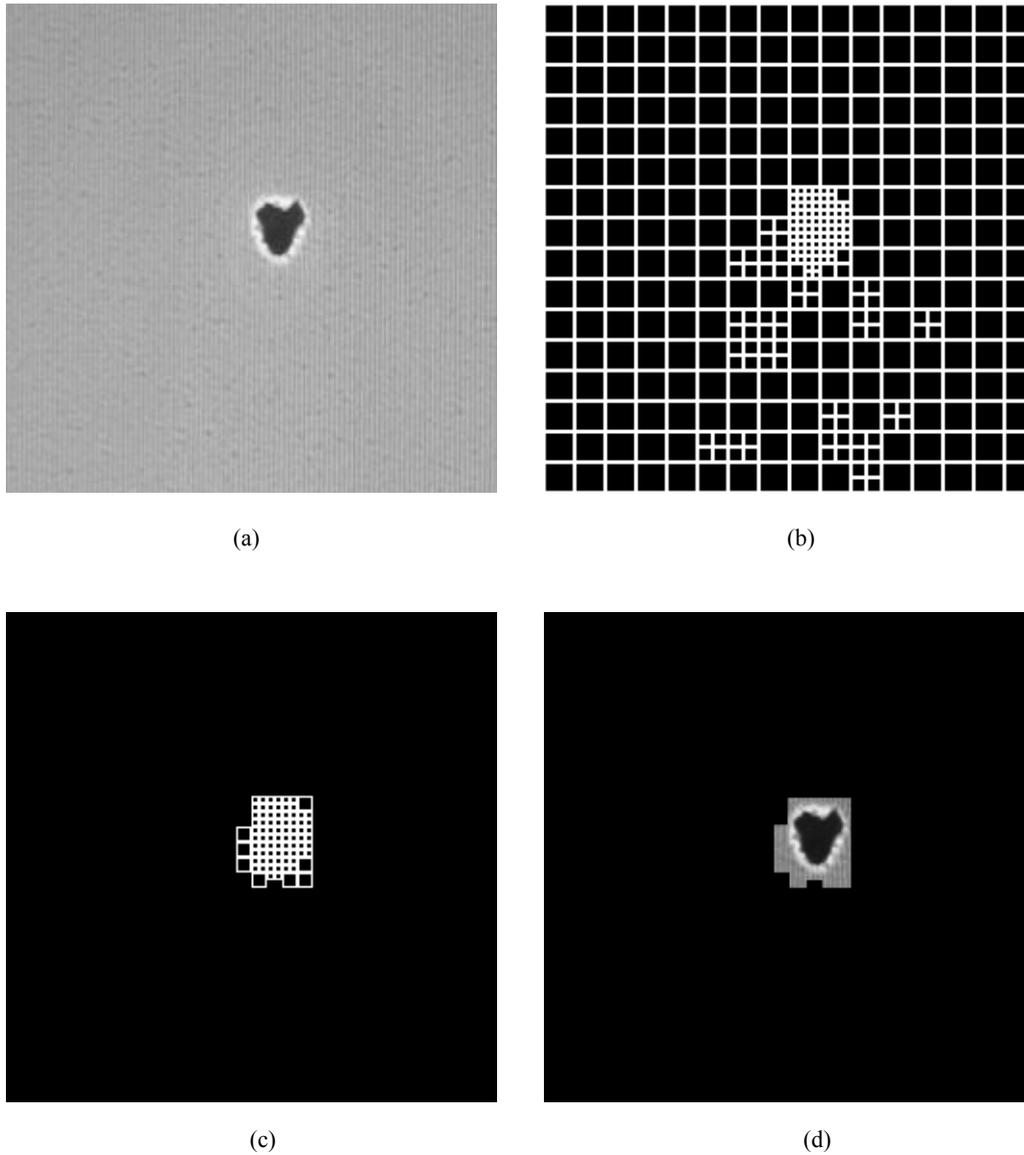


Figure 2. (a) A semiconductor image with a defect in the center of the image. (b) Quadtree partitioning as a result of the fractal encoding of the image in (a) for $L_{max} = 6$ (i.e., smallest subimages are 4 x 4) and $T = 14.0$. (c) Those subregions and their 8-neighbors in (b) that never satisfied the similarity condition. (d) Generated FARs.

4. RESULTS

To validate the NRDD approach to finding defects, several data sets of semiconductor SEM and optical images were utilized. A person familiar with semiconductor imagery viewed each of the images and, using a mouse and a graphical interface, delineated the defects manually. These results were then saved and used as a standard for evaluating the accuracy of the defect as delineated automatically by the NRDD algorithm. All of the pixels either on the boundary or included within the interior of the boundary are collectively called the *defect mask*. We call the level of similarity between the NRDD defect mask and the man-

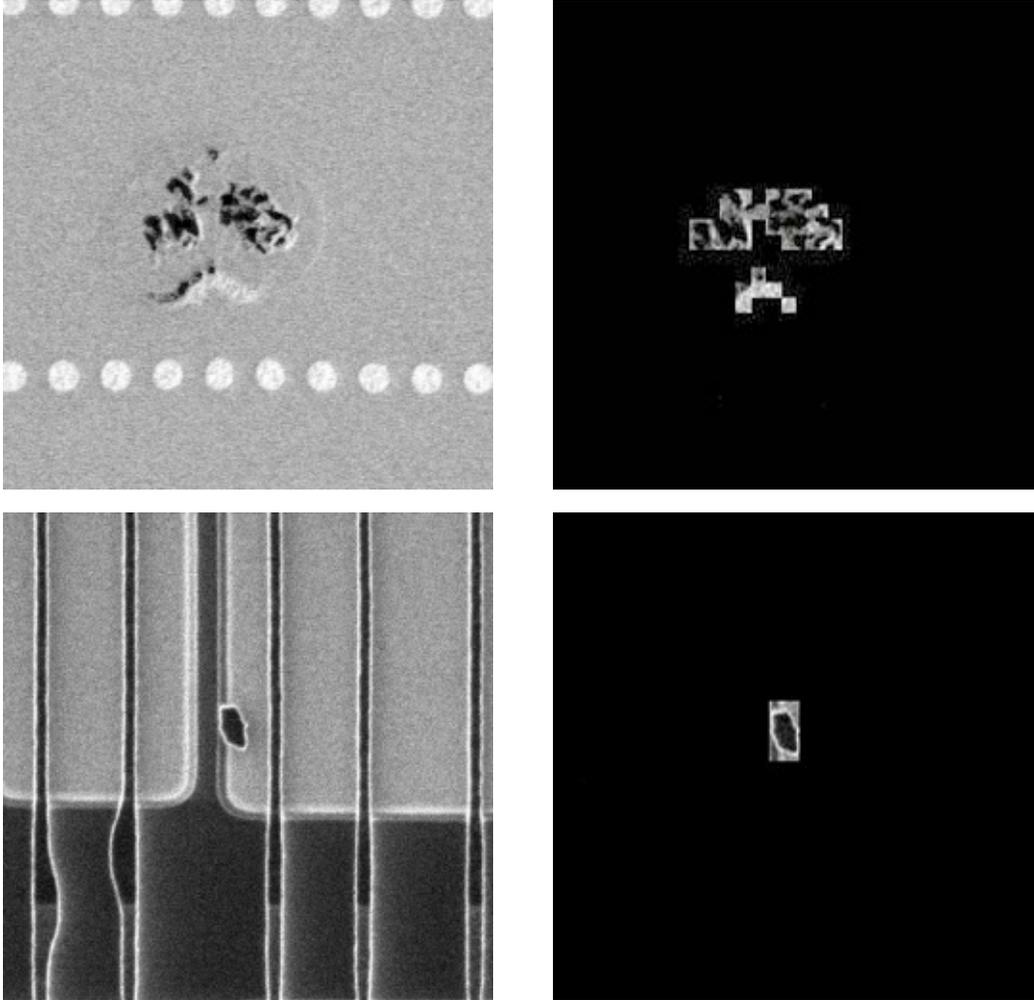


Figure 3. Original semiconductor images are shown on the left, and the FARs identified via the fractal encoding process are on the right. Note that in the first image two separate FARs are identified, but both cover different parts of the defect.

ual defect mask f_{match} , given by:

$$f_{match} = \frac{n}{(n+m)}, \quad 3$$

where n is the number of pixels in common between NRDD and manual defect masks and m is the total number of pixels that are contained in one defect mask or the other, but not common to both. A pictorial example of this is shown in Fig. 6. Based on this formula, if $f_{match} = 1$, there is a perfect correspondence of the manual and NRDD defect and if $f_{match} = 0$, there is no match between the NRDD and manual defect. With this formula the f_{match} value would be low in the case where the NRDD mask was much larger than the manual mask that it overlays because of a large m value. A small NRDD defect that is entirely overlapped by a large manual defect would result in a low f_{match} value because of the number of pixels that the NRDD defect missed.

The manual defect held up as the standard to match, is subject to interpretation and normal variation in that it was created in a

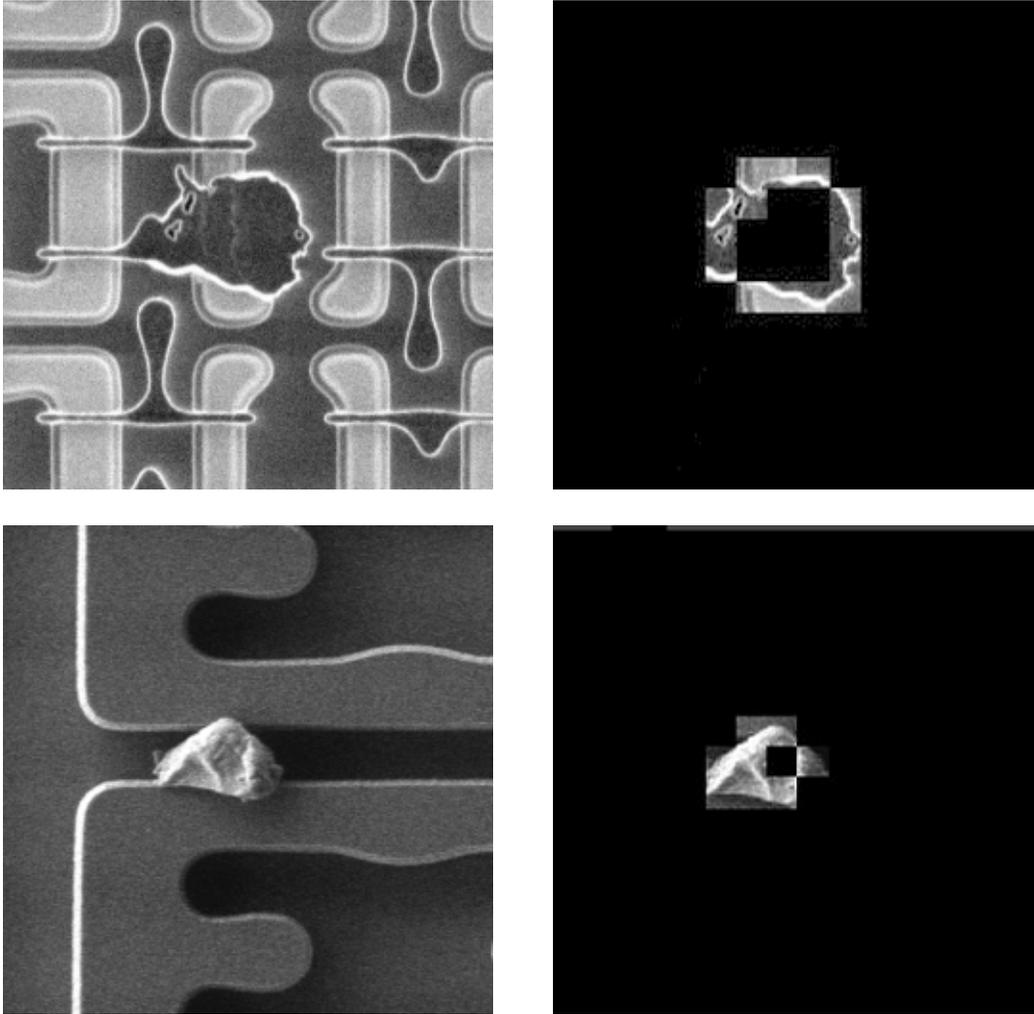


Figure 4. Original semiconductor images are shown on the left, and the FARs identified via the fractal encoding process are on the right. Note that in the top image the central portion of the defect closely resembles the background color and texture of the normal background pattern, therefore it was not covered by the FAR.

manner that yields a non-unique solution. It should be apparent that manual defects delineated by two different individuals would have a matching criterion of less than 1. Because the images used here are actual images and not simulated data, even NRDD results that appear ideal may have only had a match ratio of 0.7 - 0.8. Based on the match ratio, the NRDD defects were labeled as falling into one of three categories. The no match category corresponds to situations where there was either no overlap in the NRDD defect and the manual defect or NRDD failed to locate a defect. The poor match category with f_{match} less than 0.2 corresponds to only a small part of the NRDD and manual defect corresponding. This would also apply in situations where the NRDD and the manual defect are grossly different in size. The good match category with f_{match} greater than 0.2 applies when the NRDD defect is a reasonable approximation of the manual defect. On a highly patterned image “poor match” defect detection has frequently captured one of the irregular edges of a defect. These results are still rather impressive particularly in cases of missing or extra pattern.

The present implementation of the NRDD algorithm was developed to focus on a single defect in the field of view with an expectation that the defect of interest is most likely to be found near the center of the image. While the fractal encoding step of

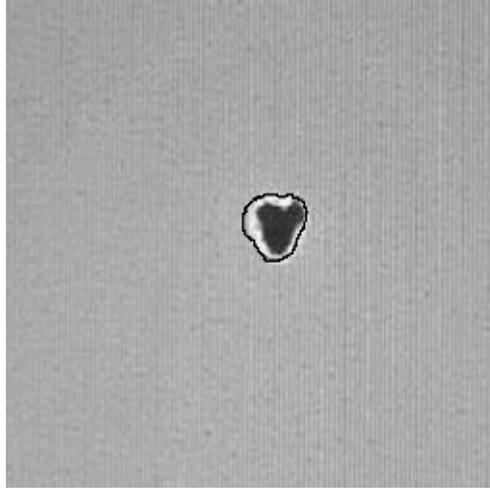


Figure 5. Result of snake-based segmentation of the defect from Fig. 2. The final contour is shown in black.

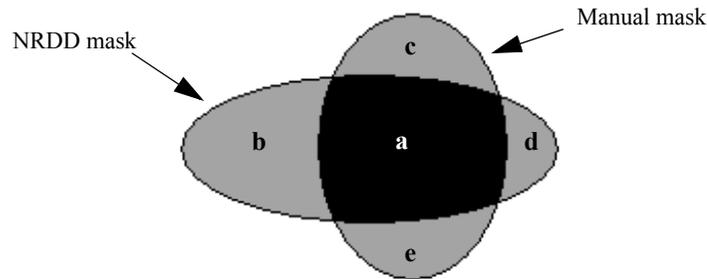


Figure 6. Contrived example of two overlapping elliptical defect masks: one generated using NRDD and one generated manually. The black area is where the two masks overlap, and the gray areas are regions that are not common to both masks. The area of each region is labelled on the figure as a, b, c, d, and e. To calculate the mask similarity, f_{match} , in this example, $n = a$ and $m = b+c+d+e$.

the algorithm may frequently identify several different focus of attention regions, for test and evaluation purposes, the algorithm selects the detected regions nearest the center for the active contour step. For the semiconductor application, this is a valid assumption. The images shown in Figures 7 through 10 illustrate the performance of this NRDD algorithm on several different types of semiconductor types.

Table 1 below details the performance of the NRDD algorithm over one optical and three SEM data sets. The average size of the defect with respect to the image is given based on the number of pixels in the manually delineated defects. The three categories are based on the value of f_{match} as given above. One of the SEM data sets on average had significantly larger defects (relative to image size) on average than the defects found in the other data sets. The SEM 2 data set was highly patterned and diverse in appearance.

Effectiveness of the NRDD algorithm is closely tied to the optimal selection of algorithm parameters for fractal encoding and for the active contours. All of the images in a data set were processed with a single set of parameters for Table 1. These parameters varied from data set to data set. A methodology for auto setting the threshold parameter, T , for the fractal encoding on a per image basis was explored and preliminary results using this methodology for two of the four data sets is shown in Table 2. Auto setting of this parameter alone has had a beneficial impact on the quality of NRDD defect detection. The details of this

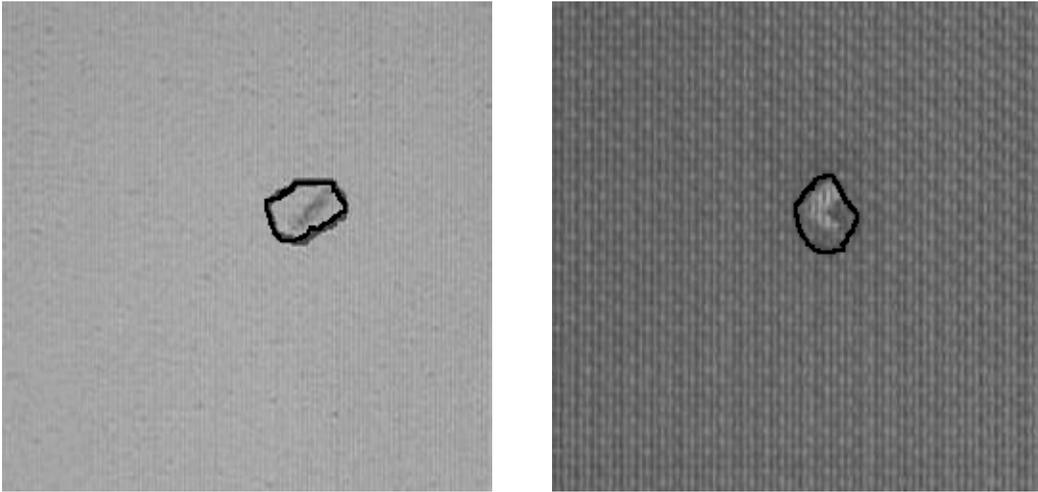


Figure 7. Example NRDD segmentations from a database of optical defect images (provided by SEMATECH). Detected defect boundary is shown in black. The f_{match} values for these examples are 0.16 (left) and 0.5 (right).

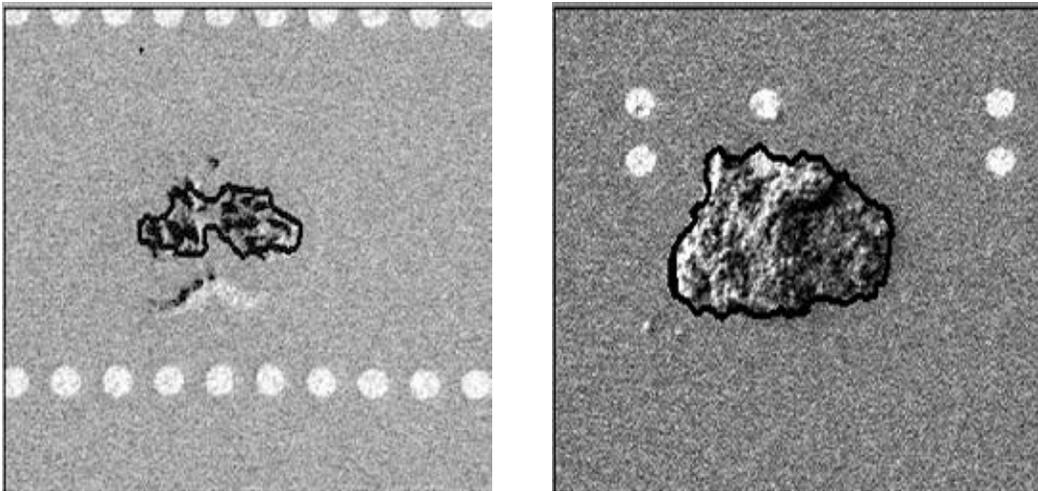


Figure 8. Example NRDD segmentations from a database of SEM (layer 1) defect images (provided by SEMATECH). Detected defect boundary is shown in black. The f_{match} values for these examples are 0.3 (left) and 0.89 (right).

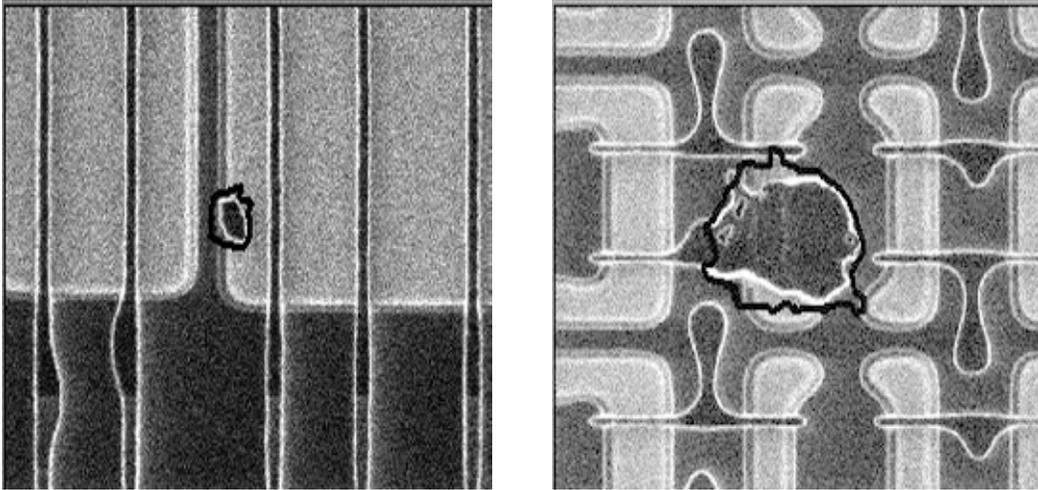


Figure 9. Example NRDD segmentations from a database of SEM (layer 2) defect images (provided by SEMATECH). Detected defect boundary is shown in black. The f_{match} values for these examples are 0.82 (left) and 0.78 (right).

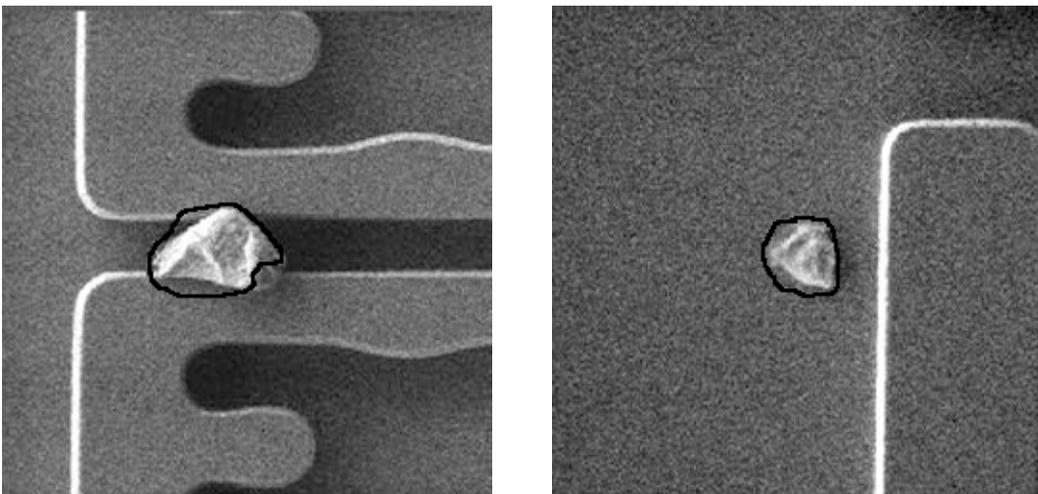


Figure 10. Example NRDD segmentations from a database of SEM (layer 3) defect images (provided by SEMATECH). Detected defect boundary is shown in black. The f_{match} values for these examples are 0.65 (left) and 0.73 (right).

auto-parameter generation procedure are not detailed in this paper, but will be published upon further validation of the technique.

Table 1: NRDD Performance with Manual Parameter Selection

Dataset identifier	Avg defect size as % image size	#images in dataset	No match	Poor match	Good match
Optical	2.15%	170	2.40%	13.50%	84.10%
SEM layer 1	7.33%	158	13.90%	7.60%	78.50%
SEM layer 2	2.82%	1136	23.90%	25.20%	50.90%
SEM layer 3	2.19%	783	21.20%	18.80%	60.00%

Table 2: NRDD Performance with Automatic Parameter Selection

Dataset identifier	Avg defect size as % image size	#images in dataset	No match	Poor match	Good match
SEM layer 1	7.33%	158	2.53%	5.70%	91.77%
SEM layer 2	2.75%	1127	15.70%	27.20%	57.10%

5. CONCLUSIONS

The NRDD algorithm presented here can be important for applications in which automatic defect detection is needed, but a reference image is unavailable. This is in general an extremely challenging task because of the variety of defects and backgrounds that may be experienced in a semiconductor manufacturing environment. The results presented in the previous section demonstrate that this approach certainly has tremendous potential for detecting a wide variety of defects on a wide variety of backgrounds. This NRDD approach is in theory and in practice applicable to multiple image modalities (e.g. SEM and optical microscope imagery). This technique does have limitations in its ability to locate small defects on very complex backgrounds (e.g. SEM layers 2 and 3), but even in these cases, the majority of defects were detected at some level. Note that improvements in performance have already been seen using the preliminary automatic parameter setting algorithm (see Table 2). The final application for this technology would dictate the level of performance required from the NRDD algorithm, and would, in turn, dictate the types of defects and layers that would be amenable to the use of the NRDD algorithm.

ACKNOWLEDGEMENTS

This work was funded by the Yield Management Tools Program at International SEMATECH, and special thanks go to Fred Lakhani who manages this program at SEMATECH.

REFERENCES

1. W. B. Jatko, M. A. Hunt, K. W. Tobin, "Nonlinear filter derived from topological image features," Proc. Int. Soc. for Optical Engineering Conf., Orlando, FL, 1990.
2. Karnowski, T.P., Tobin, K.W., Arrowood, L.F., Ferrell, R.K., Goddard, J.S., and Lakhani, F., "Field-test results of an image retrieval system for semiconductor yield learning", SPIE's 13th International Symposium on Electronic Imaging: Science and Technology, San Jose Convention Center, January 2001.
3. P. Xie, S. Guan, "A Golden-Template Self-Generating Method for Patterned Wafer Inspection," Machine Vision and Applications, vol. 12, no. 3, pp. 149-156, 2000.
4. S. S. Gleason, H. Sari-Sarraf, K. T. Hudson, and K. F. Hubner, "Higher accuracy and throughput in computer-aided screening of mammographic microcalcifications," *IEEE Medical Imaging Conf.*, 1997.

5. H. Sari-Sarraf, S. S. Gleason, and R. M. Nishikawa, "Front-End Data Reduction in Computer-Aided Diagnosis of Mammograms: A Pilot Study," Proc. SPIE Medical Imaging Conf., San Diego, CA, February 1999.
6. H. Sari-Sarraf, S. S. Gleason, K. T. Hudson, and K. F. Hubner, "A novel approach to computer-aided diagnosis of mammographic images," *3rd IEEE Workshop on Applications of Computer Vision*, 1996.
7. Y. Fisher, "Mathematical background," *Fractal Compression: Theory and Application to Digital Images*, Y. Fisher, ed., pp. 25-53, Springer Verlag, New York, 1994.
8. Y. Fisher, "Fractal image compression with quadrees," *Fractal Compression: Theory and Application to Digital Images*, Y. Fisher, ed., pp. 55-77, Springer Verlag, New York, 1994.
9. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. of Comp. Vis.*, vol. 1, pp. 321-331, 1988.
10. C. Xu and J. Prince, "Gradient vector flow: A new external force for snakes," *IEEE Proc. Conf. Comp. Vis. Patt. Recog.*, pp. 66-71, 1997.