Detection Algorithm of Particle Contamination in Reticle Images with Continuous Wavelet Transform

Chaoquan Chen and Guoping Qiu
School of Computer Science and IT
Jubilee Campus, University of Nottingham
Nottingham NG8 1BB, UK
{chaoquan.chen,qiu}@cs.nott.ac.uk

Abstract

This paper presents an inspection method of particle contamination for semiconductor reticles using continuous wavelet transform. Particle defect is considered as a singularity in the reticle image, and wavelet transform is applied to detect such an event. By taking the local maxima of wavelet transform as suspected defects, the candidate pixels under inspection reduce to a small fraction of the whole image. From the evolution of wavelet coefficients across scales, two features are extracted to identify detects among the suspected defects, Lipschitz Exponent (L.E.) and smoothing factor. With the rules of LE and smoothing factor trained from the image database, the defects can be detected with high accuracy. From the test of synthetic reticle images with defects at different size, it is concluded that the maximum scale of wavelet under which the defect is still visible can not be over about 8 times as much as the defect size. The simulation of a real reticle image and synthetic defects shows the effectiveness of the present method.

Keywords: defect detection, reticle inspection, wavelet transform, Lipschitz exponent and smoothing factor

1 Introduction

Periodic semiconductor reticle inspection is a critical component of the integrated circuit fabrication process. Even a single defect on a reticle will cause a catastrophic yield loss, since reticles (or photomasks) are used to print integrated circuits (ICs) on semiconductor wafers. Foreign material in the form of particle can impact yield significantly and must be detected repeatedly during IC production [1,2].
Most vision-based inspection techniques fall into one of two general categories: methods based on image-to-image comparison, and methods for checking generic properties and design rules. In image-to image comparison method, the image taken from the reticle is compared either with an image free of defect (so-called golden

BMVC 2001 doi:10.5244/C.15.45

image) or with the image taken from another die that is supposedly identical to the image being tested. Obviously the precise image alignment (registration) [3] is the critical step to expose the defect successfully. In the second category, the image is tested against a set of design rules or local properties, and violations are reported as defects. The local properties can be the defect size, defect texture and the gray variation around the defect etc.

Most existing defect inspection systems adopt the comparison strategy [2,4,5]. The present paper reports on a research effort in which the particle-like defect can be detected straightforward without comparison. Defects in the image are considered as singularities, and continuous wavelet transform is applied to detect such singular events. By taking the local maxima of wavelet transform across scales as suspected defects, the candidate pixels under inspection reduce to a tiny fraction of the whole image [6]. Then from the evolution of wavelet coefficients across scales, two features are extracted to identify detects, Lipschitz exponent (L.E.) and smoothing factor. With the rules of LE and smoothing factor trained from the image database, the defects can be detected with high accuracy.

The paper is organized as follows. In Section 2 we review some basic aspects of the wavelet transform. In Section 3, Lipschitz exponent and smoothing factor are introduced to describe the feature of the defects. Section 4 summarizes the algorithms and discusses the relationship between the particle size detectable and the wavelet scale applied. Section 5 presents the experiments of a real reticle image and syntactic defects.

2 Continuous Wavelet Transform

The Morlet-Grossmann definition of the continuous wavelet transform for a 1-dimensional signal $f(x)$, the space of all square integrable functions, is [7]:

$$WT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) \psi^* \left(\frac{x-b}{a} \right) dx \qquad (1)$$

Where:

- $WT(a,b)$ is the wavelet coefficient of the function $f(x)$
- $\psi(x)$ is the analyzing wavelet
- a (>0) is the scale parameter
- b is the position parameter

In Fourier space, we have:

$$F(WT(a,\omega)) = \sqrt{a} F(\omega) \Psi^*(a\omega) \qquad (2)$$

When the scale a varies, the filter $\Psi^*(a\omega)$ is only reduced or dilated while keeping the same pattern.

One of the outstanding characteristics of wavelet transform is the localization capability due to the compact support of the wavelet function in the time and frequency space [8]. With Fourier transform, there is little response to a singularity. On the contrast, it is prominent with wavelet transform. If a signal is singular at x_0, there exists a sequence

of wavelet transform maxima that converge to x_0 when the scales decreases. So we can detect all singularities from the position of the local wavelet transform maxima [9].
By continuous wavelet transform, we mean the two parameters of wavelet function, scale and position, are continuous, although in practical computation these parameters have to take discrete values. For singularity detection, the continuous wavelet transform is preferred to the discrete transform, since the singularity information across scales will be emphasized although the reconstruction information from the continuous wavelet coefficients is redundant.

3 Feature Selection

As mentioned above, one signal sharp variation such as singularities produces local maxima at different scales. Therefore, tracking the local wavelet transform maxima across scales will highlight the defects significantly. We know that the local wavelet maxima at the scale a measures the derivative of the signal smoothed at this scale, but it is not clear how to combine these different values to characterize the signal variations[10]. Moreover, for an isolated particle and a particle in a corner within an image, the wavelet transform will take the local maximum at these two positions. However, the evolution of wavelet coefficient with scales is distinguishable, because these given points consist in different detailed information at different scales. So by tracking the variation of wavelet coefficient with scales, singularities can be further classified. Two features are derived from the evolution of wavelet transform across scales: Lipschitz exponent (L.E) α and smoothing factor σ.

3.1 Lipschitz exponent

In mathematics, Lipschitz exponent is a measure to describe the local property of a function. It is defined as follow [9]. Let *n* be a positive integer and $n \leq \alpha \leq n+1$. A function *f(x)* is said to be Lipschitz α, at x_0, if and only if there exists two constants *A* and $h_0 > 0$, and a polynomial of order *n*, $P_n(x)$, such that for $h < h_0$

$$\left| f(x_0 + h) - P_n(x_0 + h) \right| \leq A \left| h \right|^{\alpha} \qquad (3)$$

In fact the polynomial $P_n(x)$ is the first *n+1* terms of the Taylor series of $f(x)$ at x_0. If $f(x)$ is differentiable at x_0, then it is Lipschitz α =1; If the uniform Lipschitz regularity α is larger, the singularity at x_0 will be more “regular”. If $f(x)$ is discontinuous but bounded in the neighborhood of x_0 (i.e. step edge), its uniform Lipschitz exponent in the neighborhood of x_0 is 0. For the 2 dimensional signal (images) similar definition can be derived from the simple extension of the above.

A function $f(x,y)$ is uniformly Lipschitz α over $(x,y) \in (a,b) \times (c,d)$ if and only if there exists a constant K >0 such that for all $(x,y) \in (a,b) \times (c,d)$, the wavelet transform satisfies [10]

$$\left|WTf(x,y)\right| \leq Ka^{\alpha} \qquad (4)$$

This shows that the Lipschitz exponent of a function can be measured from the evolution across scales of the absolute wavelet value. The method is described as follows. Firstly, perform wavelet transform across scales; secondly, locate the local maximum wavelet transform at each scale; then the following vector is obtained,

$$WT_{a_1}(f(x,y)), WT_{a_2}(f(x,y)) \cdots, WT_{a_i}(f(x,y)) \cdots \quad (5)$$

Finally by least-squares fitting, Lipschitz exponent α can be extracted from the above vector[6].

In theory, the Lipschitz exponent of ramp function is 1; For step edge it is 0, and for Dirac –1. The negative L.E means it is more singular than discontinuity. In practical computing, we obtained slightly different values from the theoretical ones. This is because we are limited by the resolution, that is, we can not compute the wavelet transform at scales finer than 1.

3.2 Smoothing Factor

Smoothing factor can be another feature together with Lipschitz exponent to describe the local sharp variations of a function. A signal is usually not singular in the neighborhood of local sharp variations. A particle defect is foreign material in the reticle maybe distinguishable from the pattern in intensity. However, in the process of producing an image a diversity of factors result in the smoothness of the variation between particles and patterns in the reticle. This makes the particle detection more challenging. We can model the smoothness variation at x_0 as a singularity convoluted with a Gaussian of variance σ^2. Here σ is called the smoothing factor, which can indicate the particle image size in a sense.

We suppose that locally, the signal $f(x,y)$ is equal to the convolution of a function $h(x,y)$, which has a singularity at (x_0, y_0), with a Gaussian $g_\sigma(x,y)$ of variance σ^2

$$f(x,y) = h(x,Y) * g_\sigma(x,y) \text{ with } g_\sigma(x,y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (6)$$

By performing wavelet transform over (6), we can prove that the wavelet transform at the scale a of a singularity smoothed by a Gaussian of variance σ^2 is equal to the wavelet transform of the non-smoothed singularity $h(x,y)$ at the scale $s_0 = \sqrt{a^2 + \sigma^2}$, i.e.,

$$WT_a f(x,y) = \frac{a}{s_0} WT_{s_0} h(x,y) \quad (7)$$

Since $\left|WT_a h(x,y)\right| \le Ks^{\alpha}$, we can obtain

$$\left|WT_a f(x,y)\right| \le Kas_0^{\alpha-1} \text{ With } s_0 = \sqrt{a^2+\sigma^2} \qquad (8)$$

The above formula shows that the parameters α,σ and K can describe different properties of the sharp variation that occurs at (x_0, y_0), which can be used as the features to classify the singularities. For example, if the signal $f(x,y)$ is amplified by K', then K becomes KK', but α and σ are not affected. On the other hand, if the signal is smoothed by a Gaussian of variance σ_0^2, then α and K remains unaffected, but σ^2 becomes $\sigma^2+\sigma_0^2$.

α,σ and K are computed as follows. If the local wavelet maxima at three scales j=1,2,3 are WT_1, WT_2 and WT_3 respectively, then (8) becomes

$$WT_1 = K(1+\sigma^2)^{\frac{\alpha-1}{2}} \qquad (9a)$$

$$WT_2 = 2K(4+\sigma^2)^{\frac{\alpha-1}{2}} \qquad (9b)$$

$$WT_3 = 3K(9+\sigma^2)^{\frac{\alpha-1}{2}} \qquad (9c)$$

By solving the above equations, we can get these three parameters. Normally the solutions obtained in this way are not stable. Actually we can have a vector as (5) for scale $a_1, a_2, \cdots, a_j, \cdots$. Then α,σ and K may be modeled using nonlinear least squares, such as Levenbrg-Marquardt method [11].

In sum, there are two strategies to extract features from the evolution of wavelet transform across scales to define the defect among the suspected defects. One is to apply Lipschitz exponent solely, and the other is to apply Lipschitz exponent and smoothing factor together. With the first method, the computation is simple, and the modeling is less accurate. Using the second method, we estimate two features to define the defect. Although the computation is complex, the defect can be detected more precisely.

4 Particle Defect Detection Procedure

We can summarize the detect procedure of particle defect in reticles as follows,

Step 1. Calculate the wavelet transform WT of the reticle images at scale a_j.

Step 2. Find the local maximum wavelet transform LMWT, and pick this pixel p_j as the defect candidate.

Step 3. Track the wavelet coefficient vector as (5), and calculate Lipschitz exponent α and smoothing factor σ with the strategies described in Section 3.

Step 4. Decide whether the pixel p_j is a true defect in terms of α or α and σ value. The rules are defined by the training of the real and synthetic reticle images with known particle defects.

There may be problems with missing true defects when the local maximum wavelet coefficient is defined as a suspected defect. As mentioned above, it has been proved that if a signal is singular at x_0, there exists a sequence of wavelet transform maxima that converges to x_0 when the scales decreases. Since each singularity consists in detailed information coving a certain range of scales, the inappropriate scale accounts for the missing true defects.
So which scale of wavelet can be chosen to define the suspected defects in the reticle is an important question. By testing a set of synthetic reticle images with particles at different size, we obtained the relationship between the detectable particle size d_p and wavelet scales a applied, as shown in Figure 1. In the synthetic images the position and the gray level intensity of particles are distributed randomly. By fitting the measured data we can see that the maximum scale s_{up} for detection is about 8.35 times as much as the detectable particle size d_p. This is reasonable because we choose Mexican hat as the wavelet. For 1 dimensional Mexican hat the compact support is around [-5,5]. So we take $s_{up} = 8d_p$. On the other hand the minimum scale s_{lo} under which the defect is visible is the particle size, i.e., $s_{lo} = d_p$, as shown in dashed line in the figure 1.

$$s_{up} = 8d_p \qquad (10a)$$

$$s_{lo} = d_p \qquad (10b)$$

In the practical inspection, (10a) and (10b) can be applied as the guide to choose wavelet scale to define the suspected defects. If we choose the scale larger than the upper bound s_{up}, the defect will not be visible. In addition, in terms of this relationship, by choosing different scales we can detect defects in different range of size. In a reticle image, the visible suspected defects under scale a are at relevant sizes.

5 Experimental Results

The following so-called two-dimensional Mexican hat is chosen as the wavelet due to its ability to diagnose singularities.

$$\psi(x, y) = (2 - x^2 - y^2)e^{-\frac{x^2+y^2}{2}} \qquad (11)$$

Figure 2 is the experimental results on a 256*256 synthetic reticle image. Figure 2(a) shows the reticle image with 6 defects, labeled as A to F. Correspondingly, the size is from 2 to 7 pixels. Using scale a from $2s_{up}$ to $7s_{up}$, the suspected defects at different size are obtained shown in Figure 2 (b) to (g). For $a = 2s_{up}$ in Figure 2(b), all the defects are visible. In Figure 2(c), Defect A and B disappear at scale $a = 3s_{up}$. Then C, D, E and F gradually disappear with the increase of scale. This is in agreement with (9a) and (9b). Actually in Figure 2(b) there are 254 defect candidates. Using the strategy 1 described in Section 3, we estimate the L.E of these defect candidates. With

the application of the Lipschitz exponent constraint [6], the candidate defect pixel reduces to 120 (out of 256*256) with the entire defect existing, as shown in Figure 2(h). Figure 3 (a) is a real reticle image with a particle defect highlighted near the center. As described above, the local maxima of wavelet coefficients are calculated across scales. The results are shown from Figure 3(b) to (g). These images show the suspected position of defects with the given defect highlighted by an enclosed white square. The defect in the image is about 4* 4 pixels. In Figure 3(a), since scale $a < s_{lo}$, the defect is not been highlighted. With the increase in scale, there are fewer local maxima. However, the defect remains visible across the scales. Up to scale a =12, there are still 121 pixels posing as possible defects. Based on these pixels, L.E. values were estimated using least-squares method from the vector of wavelet coefficients across different scales. Applying the LE constraints (-2<LE<1) across the image, the candidate pixels reduces to 7 (out of 128∗128), and the defect particle remains near the center of the image, as illustrated in Figure 3(h).

6 Conclusions and Further Works

The localization properties of continuous wavelet transform have enabled us to develop an effective method to detect particle defect in reticle images. By taking local maximum wavelet transform as the suspected defect, we can reduce the defect candidates to be a small fraction out of the whole image. Two features can be used to describe the evolution of wavelet coefficients across scales of these candidate defects, Lipschitz exponents and smoothing factor. The rule of Lipschitz exponent has been applied successfully to highlight the defect to the extent such as 7 out 128*128 in a real reticle image and 120 out of 256*256 in a synthetic image. The strategy with both LE and smooth factor will be tested in the near future.

Acknowledgement

This research is supported by an EU IST program grant Contract No. IST-1999-10334. The image used in the paper is provided by KLA-Tencor Ltd., Slough, UK.

References

[1] B. E. Dom and V. Brecher, Recent advances in the automatic inspection of integrated circuits for pattern defects, Machine Vision and Applications. 8 pp5-19, 1995

[2] M. Nikoonahad et al, Defect detection algorithm for wafer inspection based on laser scanning, IEEE Transactions on Semiconductor Manufacturing. vol. 10 No. 4 pp459-468, 1997

[3] L. G. Brown, A survey of image registration techniques. ACM Computing Surveys, vol.24 No.4 pp325-376, 1992

[4] V. Sankaran, C. M. Weber and K. W. Tobin, “Inspection in semiconductor manufacturing”, in Webster’s Encyclopaedia of Electrical and Electronic Engineering, vol. 10, pp.242 –262, Wiley & Son, NY, 1999

[5] M. Darboux, A. Falut, J. Jaxquot, and C. Doche, An Automated System for Submicrometer Defect Detection on Patterned Wafers, SPIE Optical Microlithography and Metrology for Microcircuit Fabrication, Vol. 1138 pp172-177.

[6] C. Chen, G. Qiu, Particle Defect Detection for Reticle Inspection Using Wavelet Transform, Proceeding of the 5th International Conference on Quality Control by Artificial Vision, May 21 -23, 2001, Le Creusot, France

[7] J. Starck, and Jean-Luc, Image processing and data analysis : the multiscale approach, Cambridge University Press, 1998

[8] C. Chui, Wavelets: a mathematical tools for signal analysis, SIAM monographs, 1997

[9] S. Mallat and W. Huwang, Singularity detection and processing with wavelets, IEEE Transactions on Information Theory. Vol. 38 No. 7 pp710-732, 1992

[10] S. Mallat, Wavelets for a Vision, Proceedings of the IEEE, Vol. 84 No.4 pp604-614,1996

[11] W. Press, S. Teukolsky, W. Vetterling and B. Flannery, Numerical Recipes in C, Cambridge University Press, 1992

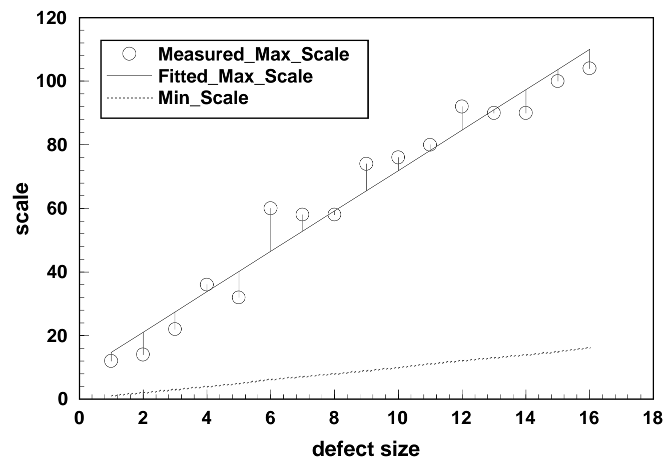
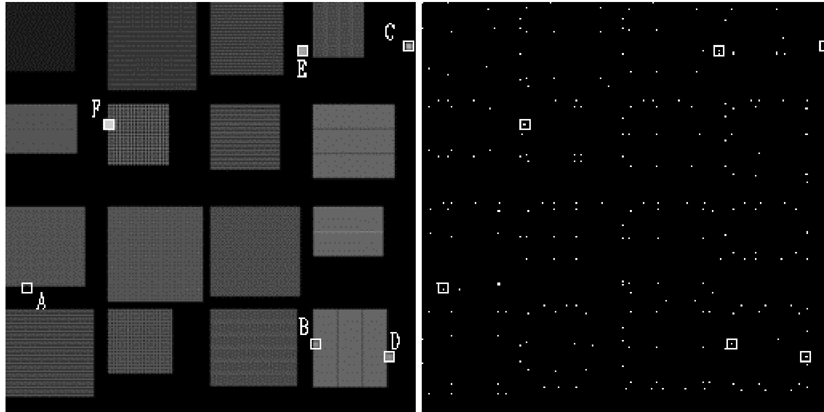
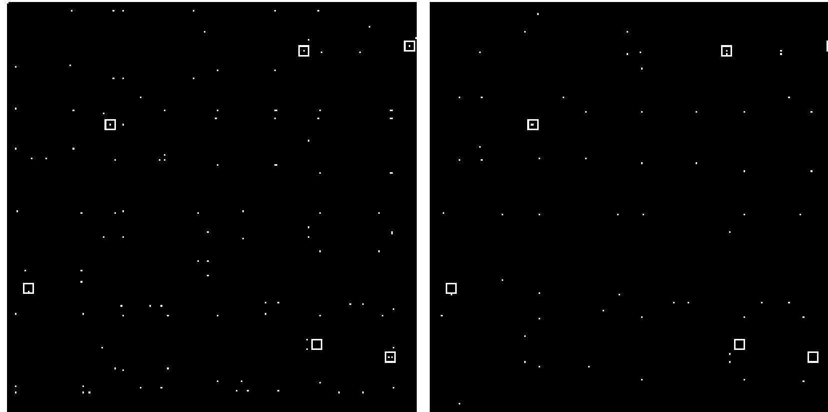


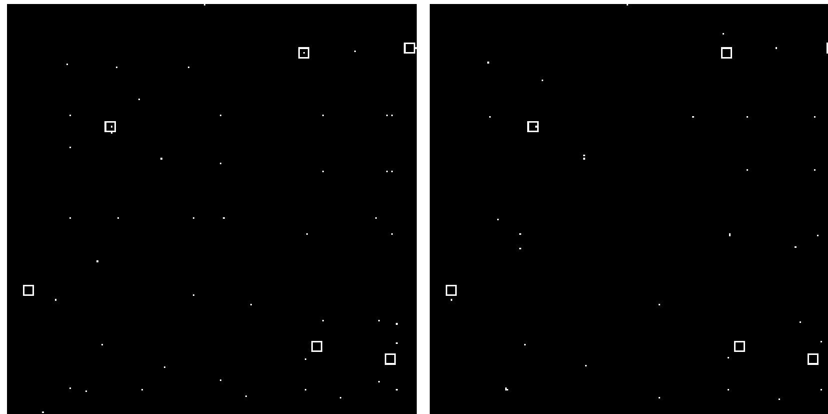
Figure. 1 detectable defect size for a wavelet scale



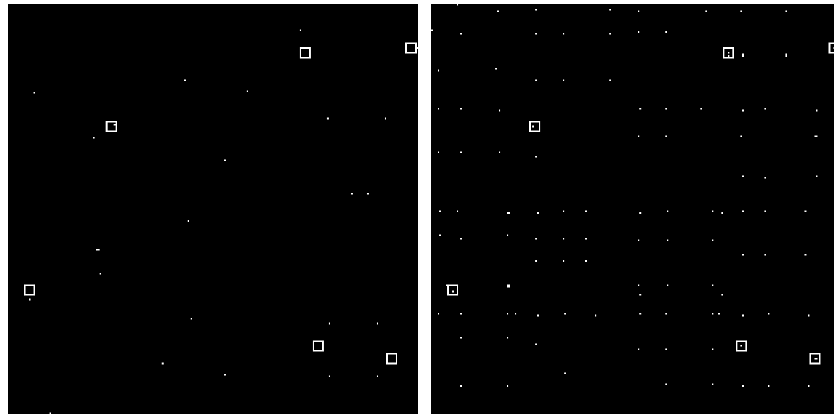
(a) Synthetic image with 6 defects: A, B, C, D, E & F (b) $a = 2s_{up}$, all visible



(c) $a = 3s_{up}$, A & B disappear (d) $a = 4s_{up}$ A, B & D disappear



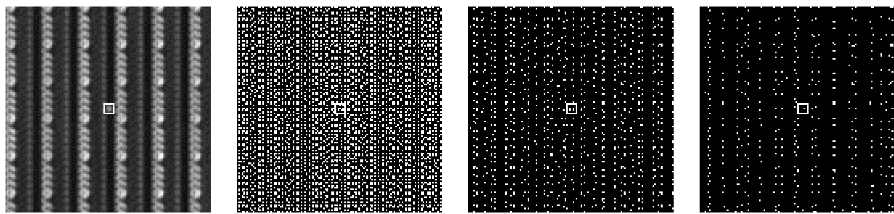
(e) $a = 5s_{up}$, A, B, C & D disappear (f) $a = 6s_{up}$ A, B, C, D & E disappear



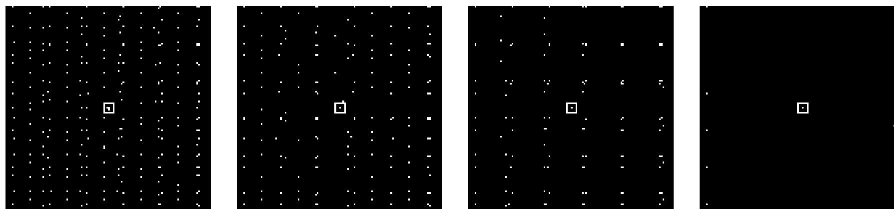
(g) $a = 7s_{up}$, A, B, C, D & E disappear (h) $a = 12$ with LE rules

A: $d_p = 2$, B: $d_p = 3$, C: $d_p = 4$, D: $d_p = 5$, E: $d_p = 6$, F: $d_p = 7$

Fig. 2 Synthetic reticle image with defects at different size



(a) Reticle image (b) $a = 2$ (c) $a = 4$ (d) $a = 6$



(e) $a = 8$ (f) $a = 10$ (g) $a = 12$ (h) $a = 12$ for $-2<\alpha<1$

Fig. 3 real defect image

(a) A real reticle image with a defect near the center (b)~(g) defect candidates defined by the local maximum wavelet coefficient across scale a (h) Defect detection specified by $-2<\alpha<1$