Multi-spot laser lock-in thermography for real-time imaging of cracks in semiconductor chips during a manufacturing process

Jinyeol Yang a, Soonkyu Hwang a, Yun-Kyu An b,*, Kyuhang Lee c, Hoon Sohn a

a Department of Civil and Environmental Engineering, KAIST, Daejeon 305-701, South Korea
b Department of Architectural Engineering, Sejong University, Seoul 143-747, South Korea
c CVI Optics Co., Ltd., 398 Seokcheon-Ro, Ohjung-Gu, Bucheon 421-809, South Korea

A R T I C L E   I N F O

Article history:
Received 16 April 2015
Received in revised form 7 September 2015
Accepted 9 September 2015
Available online 12 September 2015

Keywords:
Non-destructive testing
Multi-spot laser lock-in thermography
Real-time semiconductor chip inspection
Baseline-free crack diagnosis
Digital image processing
Diffractive optics

A B S T R A C T

This article proposes a new multi-spot laser lock-in thermography (MLLT) system for real-time imaging of cracks in semiconductor chips. The proposed MLLT system is able to inspect a semiconductor chip in real-time during its manufacturing process by simultaneously generating thermal waves on multiple points of the target semiconductor chip surface using multi-spot pulsed laser beams and measuring the corresponding thermal responses using a high-speed infrared (IR) camera. In particular, the MLLT system offers the following advantages for the semiconductor chip inspection: (1) complete non-contact, non-destructive and non-intrusive inspection, (2) real-time crack inspection with fast data acquisition and processing, (3) baseline-free crack visualization using only current-state data, making it possible to avoid false alarms caused by operational and environmental variations and (4) high detectability of cracks. To realize the MLLT system, optical components for multi-spot thermal wave generation are designed through an optical analysis and integrated with the high-speed IR camera, a close-up lens and a personal computer. The developed MLLT system is then experimentally demonstrated using actual semiconductor chips with real cracks produced during the manufacturing process. The experimental results reveal that the total inspection time including the data acquisition and processing takes less than 1 s for each semiconductor chip, and cracks in the range of 20 μm are successfully detected.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Over the last few decades, the rapid advancement in modern integrated circuit and semiconductor technology has led to the miniaturization of electronic devices. That is, the smaller and thinner semiconductor chips are developed to meet the miniaturization trend of the electronic devices. However, the semiconductor chips are inherently brittle and vulnerable to stress-induced cracks during the chip fabrication and assembly processes. For example, cracks can be initiated during wafer thinning and sawing processes. According to Islam et al. (2008), these undesired cracks can eventually cause the failure of the semiconductor chips and compromise the performance of the overall electronic devices.

To effectively inspect individual semiconductor chips during the fabrication and assembly processes, a number of non-destructive testing (NDT) techniques have been proposed. Rakotoniaina et al. (2004) proposed a contact-type ultrasound lock-in thermography technique which has high ultrasound excitation efficiency and a long penetration depth for internal defect detection. However, its contact nature for ultrasound excitation can produce stress concentration and consequent damage on brittle semiconductor chip surfaces. Recently, non-contact NDT techniques such as scanning acoustic microscopy, light scattering technique and luminescence technique have been proposed as alternatives. According to AcouLab (2012), the scanning acoustic microscopy has a relatively deep penetration depth, but it is not suitable for real-time inspection due to its long scanning time. For instance, the scanning acoustic microscopy typically takes more than 60 s to scan a single semiconductor chip of 12.8 × 8.5 mm². Moreover, the target semiconductor chip needs to be submerged into a water bath or at least covered with a water droplet for effective acoustic excitation. This wetting process often causes additional damage to the target semiconductor chips. Vision-based techniques such as the light scattering technique and the luminescence technique were proposed by Takahashi et al. (1998) and Truckle et al. (2006), respectively, as promising alternatives thanks to their non-contact nature, simple operation mechanism and short inspection time. However, not only surface contaminations or scratches but also the variation
of lighting conditions may often lead to false-positive alarms in these techniques. More recently, An et al. (2014) proposed a laser lock-in thermography (LLT) technique, which utilizes a single-spot laser source for thermal wave generation and an infrared (IR) camera for crack inspection to minimize such false alarms. However, the LLT technique is not suitable for real-time inspection of semiconductor chips due to its long data acquisition and processing time.

In this article, a new multi-spot laser lock-in thermography (MLLT) system is proposed for real-time inspection of semiconductor chips during the manufacturing process. The MLLT system is able to inspect the semiconductor chip through fully non-contact, non-destructive and non-intrusive manners. Furthermore, the newly proposed baseline-free crack visualization algorithm improves crack detectability and minimizes false alarms by alleviating the dependency on baseline data obtained from the intact condition of a target semiconductor chip, making it suitable for real-time inspection. The proposed MLLT system is composed of a continuous wave (CW) laser for non-contact thermal wave generation, a high-speed IR camera for non-contact thermal wave measurement and a personal computer for data processing and control. Specially designed multi-spot pulsed laser beams generate thermal waves at multiple spatial points of the target surface, and the corresponding 2D thermal wavefields are measured by the high-speed IR camera in the time domain. Then, the baseline-free crack visualization algorithm is developed and installed on the personal computer so that only cracks are instantaneously and automatically visualized. The performance of the proposed MLLT system is experimentally verified using commercial semiconductor chips with various cracks produced during the actual manufacturing process.

This article is organized as follows. Section 2 describes the development of the proposed MLLT system including the working principle and its optical design. Section 3 develops the baseline-free crack visualization algorithm. In Section 4, inspection test results using commercial semiconductor chips are presented. Finally, this article concludes with a summary and discussions in Section 5.

2. Development of the MLLT system for semiconductor chip inspection

This section presents the MLLT system development with its working principle and optical design. First, the size of the specific target test sample is described so that the optical components are properly designed and tailored for the given target test sample. Second, the hardware configuration and working principle of the MLLT system are described. Finally, the design of the MLLT system is verified through the optical analysis of the multi-spot pulsed laser beam excitation.

2.1. Target test sample

Semiconductor wafers are provided by Samsung Electronics Co., Ltd, and the semiconductor chips detached from these wafers are used as test samples as shown in Fig. 1. A typical semiconductor wafer contains several hundreds of semiconductor chips as shown in Fig. 1 (a). The representative semiconductor chip detached from the semiconductor wafer is then shown in Fig. 1 (b). The details of the material compounds, configurations and design parameters of the wafers and semiconductor chips are not provided in this article due to the confidentiality.

2.2. Overall configuration and working principle

Fig. 2 shows the overall schematic diagram of the proposed MLLT system composed of excitation, sensing and control units. The excitation unit consists of an arbitrary waveform generator (AWG), a laser driver, a continuous wave (CW) laser, a laser reflector, a diffractive optical element (DOE) and a plano-convex lens (PCL). The CW laser used in the system has a wavelength of 532 nm and a maximum peak power of 15 W. DOE splits a single-spot laser beam into 7 × 7 multi-spot laser beams with 0.150° of divergence, and PCL has a focal length of 1 m with the beam size smaller than 1 mm and 0.155° of divergence. The sensing and control units are comprised of the high-speed IR camera with a close-up lens and a personal computer, respectively. Here, the high-speed IR camera utilizing an uncooled microbolometer as an IR detector has a thermal time constant of 7 ms, a noise equivalent temperature difference of 30 mK, 640 × 480 pixels at a sampling rate of 50 Hz, a spectral range of 7.5–14 μm and a pixel size of 30 μm. The close-up lens has a focal length of 30 mm and a stand-off distance of 50 mm. The data acquisition and processing are controlled by LABVIEW® and MATLAB® programs installed in the control unit, respectively.

The working principle of the MLLT system is as follows. First, a modulated voltage signal generated by AWG is converted to a modulated current signal using the laser driver. The modulated current signal then activates the CW laser for generating a single-spot pulsed laser beam. The single-spot pulsed laser beam is transmitted through the laser reflector to DOE so that the single-spot pulsed laser beam is split into multi-spot pulsed laser beams. Subsequently, PCL focuses the multi-spot pulsed laser beams on the target area at the focal length of PCL. Here, the focal length of PCL is carefully designed considering the size of the target semiconductor chip and the distance between PCL and the semiconductor chip surface. The focused multi-spot pulsed laser beams generate thermal waves simultaneously at multiple spatial points of the target semiconductor chip surface. When the thermal waves reach a defect area where the thermo-physical properties abruptly change compared to the surrounding intact area, the patterns of thermal wave propagation also vary. Such variation of the thermal wave propagation is captured by the high-speed IR camera with the close-up lens in the sensing unit. The measured thermal responses are then transmitted to and stored on the personal computer. Finally, the measured data are automatically processed, and cracks are subsequently visualized using the baseline-free crack visualization algorithm installed on the personal computer.

2.3. Design of the optical component for multi-spot pulsed laser beam excitation

Within the excitation unit of the MLLT system, DOE is used for splitting a single-spot pulsed laser beam into k × k multi-spot pulsed laser beams. The employment of the multi-spot excitation expedites the inspection process and speed up the subsequent data processing by a factor of k² compared to the single-spot laser beam scanning system. For successful crack visualization, the optical components are carefully designed so that (1) each individual laser beam is well focused and localized without any divergence, (2) multi-spot pulsed laser beams are equally spaced, and (3) their intensities are uniformly distributed over the target area of interest.

Fig. 3 describes the working principle and detailed design parameters of the optical components such as DOE and PCL. The centers of DOE and PCL are both aligned with the optical axis. The single-spot pulsed laser beam is split into multi-spot pulsed laser beams using DOE. Here, the forward laser beams through DOE diverge with a specific forward distribution angle (D), and DOE is tilted by an angle of θ to avoid the zero-order at the center of the projected target surface. Note that, with the zero-order phenomenon, the energy of the laser beam at the center position falls into zero. The multi-spot pulsed laser beams are then transmitted through PCL with a backward distribution angle (D′) and focused at a focal length (L). Here, the beam diameter (d) of each laser beam
Fig. 1. Semiconductor chip samples: (a) a semiconductor wafer (φ 300 mm) and (b) a representative semiconductor chip unit (12.8 mm × 8.5 mm × 35 μm) detached from the semiconductor wafer.

Fig. 2. Schematic diagram of the proposed MLLT system for the semiconductor chip inspection: the control unit sends out control and trigger signals to the excitation unit to inject multi-spot pulsed laser beams into the target semiconductor chip for thermal wave generation, and the sensing unit simultaneously measures the corresponding thermal responses. Then, the measured thermal responses are subsequently transmitted to, stored at, and processed in the control unit.

Fig. 3. Design of the optical component for multi-spot pulsed laser beam excitation.

Fig. 4. The MTF chart for optical analysis: (a) field angle description and (b) MTF chart.
and the multi-spot spacing ($i$) between two adjacent laser beams are controlled by adjusting the distance between DOE and PCL ($i$). The main objective of the optical analysis is to determine the $l$ values by controlling design parameters such as $L$, $D$ and $D'$. The detailed design procedures are described below:

2.3.1. Determination of $d$ and $i$

Teng et al. (2011) reported that if $i$ is too small, two adjacent laser beams can overlap each other and create overheated regions. On the other hand, if $i$ becomes too large, the cracks within underheated regions can be missed. In the same manner, the value of $d$ should also be carefully tailored. According to Harp et al. (2008), a highly focused laser beam with a small $d$ value can cause ablation of the target surface while a large $d$ value may result in no thermal wave generation. Indeed, the selections of $i$ and $d$ heavily depend on the material properties of the target semiconductor chips. Based on preliminary tests using the single-spot laser beam scanning system, the $i$ values in the horizontal and vertical directions are set to 2.3 mm and 1.8 mm, respectively, and $d$ is set to 1 mm.

2.3.2. Selection of DOE and PCL

Once the target $d$ and $i$ values are determined, DOE, PCL and the optical values of $L$, $D$ and $D'$ are consequently designed. In this study, a commercial DOE, which generates a $7 \times 7$ multi-spot pulsed laser beams with $D$ of 0.095 mm, and PCL with $L$ of 1000 mm and $D'$ of 0.088 mm are used. The specifications of the optical components used in this study are summarized in Table 1. Then, $l$ is calculated based on the following equation:

$$i=ID + LD'$$

(1)

2.4. Optical analysis for the verification of the optical component design

The design of the optical components is verified using optical analysis software such as LightTools® and Code V®. First, the modulation transfer function (MTF) chart, which is the function of a spatial frequency ($\xi$) defined as $1/i$, is used to estimate the spatial distortion of the multi-spot pulsed laser beams. Second, the illumination simulation is performed to check the uniformity of the laser beam intensity over the target surface. The specifications of the CW laser used for the optical analysis are summarized in Table 2.

In geometric optics, the alteration of the monochromatic laser beam on the target surface due to diffraction or scattering phenomena causes the spatial distortion. The intensity profile shown in Fig. 5 illustrates that the desired values of $d = 1$, $i_i = 2.3$ mm and $i_v = 1.8$ mm are successfully achieved using the selected optical parameters, and the target surface is evenly covered by the simulated multi-spot pulsed laser beams.

3. Baseline-free crack visualization algorithm

This section explains how only crack is extracted and visualized from the measured raw thermal images without relying on the baseline data obtained from the intact condition of the target chip. When the multi-spot pulsed laser beams impinge to the target chip surface, thermal waves are generated and propagate along the chip surface. John and John (2008) investigated that if the propagating thermal waves encounter a crack, the thermal waves are blocked due to the thermal conductivity difference at the crack interface. The majority of thermal waves cannot pass through the crack interface, and the thermal wave energy is cumulated at the crack interface due to the thermal conductivity difference. The cumulated thermal energy can be captured and accentuated using the IR camera even when the crack width is smaller than the spatial resolution of the IR camera. However, the strong thermal waves near the excitation laser beams can mask the appearance of the thermal energy accumulation in the vicinity of the crack, and the surface patterns of the semiconductor chip can produce false indication of cracks. To extract and isolate only crack-induced thermal energy
accumulation, a baseline-free crack visualization algorithm is developed using a holder exponent analysis in this section. Such baseline-free crack detection process is able to provide instantaneous and automated diagnostic results to users and to minimize false-alarms due to operational and environmental variations, making it suitable for real-time inspection.

3.1. Computation of a lock-in amplitude image

Once raw thermal images are collected using the MLLT system, a lock-in amplitude image proposed by An et al. (2014) is computed to remove the thermal components uncorrelated with the controlled laser excitation and to extract the thermal wave components locked-in only with the excitation laser. Busse et al. (1992) reported that the lock-in amplitude image is typically sensitive to surface defects rather than subsurface defects, while the lock-in phase image is more effective to detect subsurface defects than the surface ones. In this study, since we focused on the detection of surface micro-cracks on semiconductor chips, the lock-in amplitude image is only considered. When the chip surface is exposed to a modulated laser beam as shown in Fig. 6(a), the surface undergoes heating and cooling processes as shown in Fig. 6(b). The lock-in amplitude value at a specific spatial point \((x, y)\) is computed from the raw thermal responses measured in the time domain,

\[
A(x, y, T) = \sqrt{\left\{ R(x, y, 0) - R(x, y, T) \right\}^2 + \left\{ R(x, y, \tau) - R(x, y, T) \right\}^2}
\]  

(3)

where \(R(x, y, t)\) is the thermal response at a spatial point \((x, y)\) and at a specific time \(t\). \(\tau\) and \(T\) denote the pulse width and the period of the pulse excitation, respectively. Eq. (3) computes the thermal gradient caused only by the modulated laser beam. That is, the thermal gradient synchronized only with the excitation laser beam appears in \(A(x, y, T)\), while other thermal gradients, which are not synchronized with the excitation laser beam, do not appear in \(A(x, y, T)\). Once the lock-in amplitude values are calculated for the entire area of interest, the corresponding lock-in amplitude image can be constructed as shown in Fig. 6(d).

3.2. Crack visualization based on a holder exponent analysis

Once the lock-in amplitude image is constructed, the crack feature is extracted based on the holder exponent analysis. Fig. 7 shows the working principle of the holder exponent analysis to find the spatial discontinuity caused by the crack-induced thermal energy accumulation. First, all pixel values of the row-wise discontinuity image, \(X\), are initialized to zero. Second, a wavelet transform of the nth row of the lock-in amplitude image, \(X_n\), is computed, and the absolute value of a resultant coefficient is obtained using the following equations:

\[
|Wx_n(x, s)| = \left| \int_{-\infty}^{\infty} X_n(u) \left( \frac{1}{\sqrt{s}} \right)^{\Psi^*} \left\{ \frac{|u-x|}{s} \right\} du \right|
\]  

(4)

where \(Wx_n(x, s)\) is the wavelet transform coefficient at spatial translation \(x\) and scale \(s\). \(u\) is a spatial point, and \(\Psi^*\) is complex conjugate of \(\Psi\). Here, a Haar wavelet basis function defined below is used for \(\Psi\):

\[
\Psi(u) = \begin{cases} 
1 & 0 \leq u \leq \frac{1}{2}, \\
-1 & 1/2 \leq u \leq 1, \\
0 & \text{otherwise}
\end{cases}
\]  

(5)

The holder exponent value of \(X_n\) at a specific spatial point \(x\) is simply calculated from the slope of the wavelet coefficient along the \(x\)-axis using Eq. (4). The maximum holder exponent value among all the pixel points of the nth row is registered to the corresponding pixel point in \(X_n\). The computation of the maximum holder exponent value is repeated for all rows in the lock-in amplitude image, completing \(X\). Similarly, the calculation process is repeated in the column-wise direction for obtaining a column-wise discontinuity image \(Y\). Finally, the \(ij\)-th pixel value of a discontinuity image, \(Z\), is obtained from \(X_{ji}\) and \(Y_{ij}\) as follows:

\[
Z_{ij} = \sqrt{(X_{ji})^2 + (Y_{ij})^2}
\]  

(6)

where \(X_{ji}\) and \(Y_{ij}\) denote the \(ij\)-th entity of \(X\) and \(Y\), respectively. Since the maximum holder exponent value displayed in \(Z\) reveals the local spatial discontinuity of the thermal wave propagation, the crack-induced discontinuity can be extracted and visualized.

4. Experimental investigation

The proposed MLLT system with the baseline-free crack visualization algorithm is experimentally verified using the actual semiconductor chips described in Section 2.2. The tests are performed using one intact chip and two defect chips. Fig. 8(a) and (b) show the two chips with vertical and horizontal cracks, which are produced during the wafer back-grinding process, respectively. The widths of the vertical and horizontal cracks are estimated to be approximately 20 \(\mu\)m and 4 \(\mu\)m, respectively, under microscopic observation.

Fig. 9 shows the experimental setup of the proposed MLLT system with a semiconductor chip sample. First, the CW laser (TMA-532-15T, TMA Co., Ltd.) generates a single-spot CW laser beam with a wavelength of 532 nm, and AWG generates (Agilent 33220A) a single pulse signal with a pulse width of 100 ms and 0.5 s period. Then, the single-spot CW laser beam is modulated to the single-spot pulsed laser beam through AWG and a laser driver. Here, the power of the single-spot pulsed laser beam is set to 483 mW, which is equivalent to the laser power intensity of 12.81 mW/mm². With the given laser power intensity, the surface temperature of the semiconductor chip rises only to 35.17 °C from the room temperature. The temperature of semiconductor chips can often escalate up to 175 °C during the manufacturing processes such as die attaching, wire bonding and molding. Note that the laser power level and the subsequent temperature rise due to the laser beam do not compromise the integrity and performance of the semiconductor chips. Next, the single-spot pulsed laser beam is transformed into the multi-spot pulsed laser beams through DOE and PCL. Here, PCL is 30 mm apart from DOE. The multi-spot pulsed laser beams are radiated onto the desired inspection region to generate thermal
waves. The corresponding thermal responses are measured in the time domain by the high-speed IR camera (VarioCam® hr, InfraTec) triggered by AWG. The high-speed IR camera is 50 mm apart from the target semiconductor chip and acquires raw thermal images.
Fig. 9. Experimental setup of the proposed MLLT system for crack detection in a semiconductor chip.

Fig. 10. The representative raw thermal images acquired from the horizontally cracked semiconductor chip: (a) during the heating process and (b) after finishing the cooling process.

Fig. 11. The lock-in amplitude images obtained from the raw thermal images: (a) intact semiconductor chip, (b) vertically cracked semiconductor chip and (c) horizontally cracked semiconductor chip.

Fig. 12. The discontinuity images obtained from the lock-in amplitude images shown in Fig. 11: (a) intact semiconductor chip, (b) vertically cracked semiconductor chip and (c) horizontally cracked semiconductor chip.

with a sampling rate of 50 Hz. Note that the data acquisition time of the proposed MLLT system is 25 times shorter than that of the single-spot laser beam scanning system. Here, the data collection takes 0.5 s.

Fig. 10 shows representative raw thermal images obtained from the semiconductor chip with the horizontal crack. The surface pattern of the semiconductor chip can be observed during the heating process as shown in Fig. 10(a). Even after finishing the
cooling process, the surface pattern more clearly appears in the raw thermal image as shown in Fig. 10(b), meaning that surface pattern-reflected thermal responses are time-invariant. However, the computed lock-in amplitude image clarifies crack-induced temperature gradient due to the thermal blocking phenomenon by successfully removing the time-invariant surface pattern effect.

Once the raw thermal images are obtained, the lock-in amplitude images for the three samples are computed using Equation (2), as shown in Fig. 11. The surface patterns of the semiconductor chips are successfully removed. In Fig. 11(a), only temperature increase due to the multi-spot excitation is visualized. On the other hand, Fig. 11(b) and (c) clearly show the thermal blocking phenomena caused by the vertical and horizontal cracks, respectively.

Next, the lock-in amplitude images are subsequently processed using the holder exponent analysis described in, (Eqs. (4)–(6). Fig. 12(a)–(c) show the discontinuity images obtained from the lock-in amplitude images in Figs. 11 (a), (b) and (c), respectively. In Fig. 12(a), the laser-generated thermal images are eliminated, and only occasional salt and pepper noises are observed. On the other hand, the thermal blockings caused by the vertical and horizontal cracks are clearly visualized in Fig. 12(b) and (c), respectively. Note that the data processing time is less than 0.5 s with MATLAB® program run on the 3.60 GHz processor (Intel® CoreTM, i7-4790). Overall, the entire inspection time for each semiconductor chip is less than 1 s.

5. Conclusion

In this article, a multi-spot laser lock-in thermography (MLLT) system is developed so that thermal waves on a semiconductor chip surface can be generated at multiple locations using multi-spot pulse laser beams and the corresponding thermal responses can be measured by a high-speed infrared camera. Then, a baseline-free crack visualization algorithm is developed and applied to the measured thermal images for automated and instantaneous detection of micro-cracks. Experimental verification performed on three actual semiconductor chips, one intact and two with vertical and horizontal cracks, reveals that (1) micro-cracks in the range of 4–20 µm width can be successfully detected, (2) surface patterns in the semiconductor chips are successfully removed from the thermal images using the proposed crack visualization algorithm and do not cause any false indication of cracks, (3) the entire inspection process including data acquisition and processing is fully automated and takes less than 1 s for each chip, making it attractive for real-time inspection, and (4) the employed multi-spot laser beams do not cause any damage to the semiconductor chip, achieving fully non-contact, non-destructive, and non-intrusive inspection. It is envisioned that the MLLT system can be further advanced as a stand-alone system in the semiconductor manufacturing facilities for the real-time inspection of various wafer levels as well as individual chip levels.

Acknowledgements

This work was supported by Climate Change Research Hub (CRH) of Korea Advanced Institute of Science and Technology (KAIST) and Samsung Electronics Co., Ltd.

References