

# In-line imaging and recognition of flip chip fabrication defects by real-time photoacoustic remote sensing system

Jijing Chen<sup>a,1</sup>, Kaixuan Ding<sup>a,1</sup>, Yihan Pi<sup>a,1</sup>, Shoujun Zhang<sup>a</sup>, Jiao Li<sup>a,d,e,\*</sup>, Zhen Tian<sup>a,b,c,e,\*</sup>

<sup>a</sup> College of Precision Instrument and Optoelectronics Engineering, Tianjin University, Tianjin 300072, China

<sup>b</sup> Center for Terahertz Waves, Tianjin University, Tianjin 300072, China

<sup>c</sup> Key Laboratory of Optoelectronics Information and Technology (Ministry of Education), Tianjin 300072, China

<sup>d</sup> Tianjin Key Laboratory of Biomedical Detecting Techniques and Instruments, Tianjin University, Tianjin 300072, China

<sup>e</sup> Georgia Tech Shenzhen Institute (GTSI), Tianjin University, Shen Zhen 518067, China

## ARTICLE INFO

### Keywords:

Photoacoustic  
In-line monitoring  
Parallel processing  
Deep learning

## ABSTRACT

Microscopic defects in flip chips, originating from manufacturing, significantly affect performance and longevity. Post-fabrication sampling methods ensure product functionality but lack in-line defect monitoring to enhance chip yield and lifespan in real-time. This study introduces a photoacoustic remote sensing (PARS) system for in-line imaging and defect recognition during flip-chip fabrication. We first propose a real-time PARS imaging method based on continuous acquisition combined with parallel processing image reconstruction to achieve real-time imaging during the scanning of flip-chip samples, reducing reconstruction time from an average of approximately 1134 ms to 38 ms. Subsequently, we propose improved YOLOv7 with space-to-depth block (IYOLOv7-SPD), an enhanced deep learning defect recognition method, for accurate in-line recognition and localization of microscopic defects during the PARS real-time imaging process. The experimental results validate the viability of the proposed system for enhancing the lifespan and yield of flip-chip products in chip manufacturing facilities.

## 1. Introduction

Over recent decades, flip-chip technology has significantly advanced the performance of electronic devices. The flip-chip fabrication process, involving physical vapor deposition, chip bumping, flipping, and filling, may introduce defects, including delamination, cracks, scratches, and solder bump voids. These defects typically span from micrometers to sub-millimeters and can impact chip lifespan and product yield. At present, chip manufacturers rely on post-fabrication sampling methods like electrical tests for chip performance evaluation [1–3]. However, these methods lack the real-time monitoring of the defects in flip-chip manufacturing, constraining their potential to optimize product yield and lifespan. Furthermore, with the escalation of chip integration levels, defects of micrometer dimensions, termed as microscopic defects, can exert more pronounced effects on chip lifespan and yield [4]. Hence, the imperative development of in-line, non-destructive, and precise defect monitoring technology is essential. The technology enables real-time imaging and precise identification of microscopic defects of flip chips

during critical fabrication stages, facilitating the early removal of chips with fabrication flaws and thereby improving product yield and production efficiency.

Presently, state-of-the-art flip-chip manufacturing procedures demand in-line non-destructive defect monitoring technologies that adhere to contamination-free testing standards, offer real-time imaging capabilities, possess a high spatial resolution, and can seamlessly integrate with production lines [5–7]. Until now, various technologies have been explored in the field of in-line non-destructive defect monitoring for flip-chip applications. Ultrasonic technology has been employed to assess the bonding quality of layered materials or detect minor defects within flip-chip structures [8–11]. The surface acoustic wave (SAW) method has been employed for non-destructive evaluation of mechanical properties, residual stress, and subsurface damage in ground silicon wafers [12,13]. X-ray computed tomography (X-CT) technology can offer sub-micron 3D imaging of integrated circuit chips [14–16]. Pulsed phase thermography (PPT), as a promising in-line non-destructive defect monitoring technology, has the potential to characterize issues like

\* Corresponding authors at: College of Precision Instrument and Optoelectronics Engineering, Tianjin University, Tianjin 300072, China.

E-mail addresses: [jiaoli@tju.edu.cn](mailto:jiaoli@tju.edu.cn) (J. Li), [tianzhen@tju.edu.cn](mailto:tianzhen@tju.edu.cn) (Z. Tian).

<sup>1</sup> These authors contributed equally to this work.

missing bumps in high-density packages [17,18]. However, the discussed technologies encounter challenges in in-line non-destructive monitoring. Ultrasonic techniques necessitate liquid coupling agents that may contaminate chips and cause metal structure rust on the production line. SAW technology typically requires direct contact between the transducer and the sample, and trade-offs may be necessary to achieve high-resolution defect detection and deep probing. X-CT requires time-consuming data processing and carries the risk of ionizing radiation. PPT entails a complex and time-consuming data processing procedure when identifying defects in chip metal structures. Therefore, further development is essential to effectively apply in-line non-destructive defect monitoring technologies in flip-chip production lines.

Photoacoustic (PA) imaging, which combines the advantages of high-resolution optical imaging and the deep penetration depth of ultrasound imaging, has made significant strides in the field of non-destructive testing [19–21]. Unfortunately, traditional photoacoustic imaging devices using ultrasound transducers alongside liquid coupling agents significantly hinder the application of this technology for in-line monitoring in chip production lines. To overcome the limitations of traditional PA technology requiring liquid couplants, all-optical non-contact techniques, such as laser ultrasound (LU), have been employed for non-destructive testing in various applications [22–24]. Typically, the resolution of this approach remains below sub-millimeter levels, facilitating non-destructive detection of chip defects at this dimensional scale [25,26].

In 2017, another all-optical non-contact imaging technique with micrometer-level resolution was introduced, termed photoacoustic remote sensing (PARS) microscopy [27]. Grounded in the theory of elasto-optical, this technique, characterized by its high depth-to-resolution ratio and non-contact imaging capabilities, combined with various rapid imaging strategies, has demonstrated its capacity for wide field of view (FOV) and high-speed imaging [28–31]. Furthermore, PARS has preliminarily achieved non-destructive testing with high resolution and a wide FOV in flip-chip samples [32]. However, to meet the demand for in-line non-destructive and precise defect monitoring of flip-chips during the fabrication process, further research on PARS technology is imperative. On one hand, to timely grasp the structural information of samples during the fabrication process, real-time imaging of flip-chips is required by the PARS technique. On the other hand, due to the small proportion of pixels occupied by microscopic defects in PARS images and the limited information they carry, there is a propensity for missed detections during the recognition process. Therefore, to achieve accurate in-line recognition of microscopic defects in samples, PARS technology needs to incorporate convolutional neural networks with higher object recognition precision for intelligent identification in-line.

Herein, we propose a real-time PARS system designed for in-line non-destructive defect monitoring of flip-chips during the fabrication process. Building on our prior research [32], our work introduces a real-time imaging method of PARS based on continuous acquisition combined with parallel processing image reconstruction, facilitating live imaging in fabricated flip-chip samples. During the real-time imaging process, we introduce a defect recognition method named improved YOLOv7 with space-to-depth block (YOLOv7-SPD) based on deep learning, designed for precise in-line recognition of microscopic defects within flip-chip samples. With its non-contact rapid imaging capabilities and superior micro-defect recognition, the PARS-based in-line non-destructive defect monitoring system exhibits promising potential for seamless integration into flip-chip production lines. Consequently, it stands as a robust tool for quality assurance and process optimization within flip-chip manufacturing.

## 2. Mechanism of imaging in flip-chip samples through elasto-optical effect

In this section, we exclusively focus on the elasto-optical mechanism

at the interface between undoped silicon and metal. In PARS, the fundamental description of the elasto-optic effect posits that the initial photoacoustic pressure generated by the optical energy absorbed by the medium modulates the intrinsic refractive index, as given by the following equation:

$$\delta n_{\text{material}}(z, t) = \frac{\varepsilon n_0^3 p_{\text{material}}(z, t)}{2\rho v_s^2} \quad (1)$$

here,  $p_{\text{material}}(z, t)$  represents the photoacoustic initial pressure generated due to the medium absorbing the laser pulse energy, its magnitude being related to the medium's absorption coefficient  $\mu_{a\text{-material}}$ .  $\delta n_{\text{material}}(z, t)$  denotes the modulation of the photoacoustic initial pressure  $p_{\text{material}}(z, t)$  on the intrinsic refractive index  $n_0$  of the medium (prior to excitation).  $\varepsilon$  signifies the local elasto-optic coefficient,  $\rho$  stands for the local mass density, and  $v_s$  represents the speed of sound within the medium. At this time, the continuous probe beam co-focusing with the excitation beam induces changes in reflectivity at the sample boundaries, serving as the measured physical parameter in our experiments:

$$\Delta R(n_1, n_2, \delta n_1) = \left| \frac{n_1 + \delta n_1 - n_2}{n_1 + \delta n_1 + n_2} \right|^2 - \left| \frac{n_1 - n_2}{n_1 + n_2} \right|^2 \quad (2)$$

here,  $n_1$  denotes the intrinsic refractive index of the absorbing medium, while  $n_2$  represents the intrinsic refractive index of the non-absorbing ambient medium. For the flip-chip samples composed of silicon-aluminum investigated in this manuscript, when employing 1064 nm pulsed excitation light, the material's absorption of the excitation light is related as follows:  $\mu_{a\text{-Al}}(1.0927 \times 10^6 \text{ cm}^{-1}) \gg \mu_{a\text{-Si}}(11.1 \text{ cm}^{-1})$ , satisfying  $\frac{\mu_{a\text{-Al}}}{\mu_{a\text{-Si}}} \sim 10^6$ . In this context, with silicon wafer coated with a complete aluminum film, aluminum can be regarded as the absorbing medium and silicon as the ambient medium. When there is a peeling of the metal film in the sample, which corresponds to a delamination type of defect [33,34], silicon is considered as the absorbing medium and air as the ambient medium.

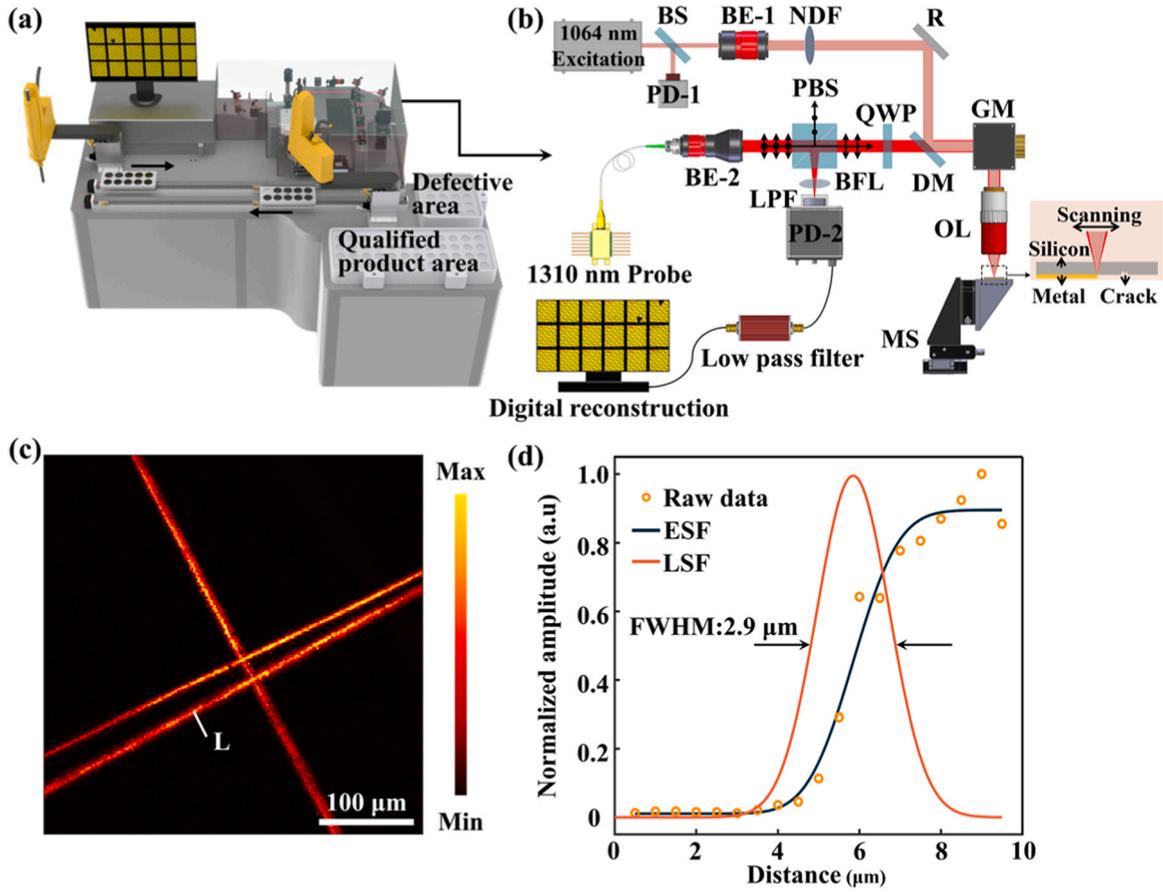
The PARS signals corresponding to the silicon wafer with metal film and delamination regions in the flip-chip samples we used are shown in Fig. S1(a). It can be observed that the region with metal film produces a positive-slope PARS signal, while the delamination area generates a negative-slope PARS signal. Utilizing the distinct slope characteristics of the PARS signals can aid in defect recognition during the flip-chip fabrication process. For instance, a negative-slope PARS signal, indicative of the silicon-air interface, may suggest instances of delamination or missing solder bumps within the substrate. Consequently, the advancement of real-time imaging and in-line defect recognition methods can further confirm whether suspicious regions contain similar defects.

On the other hand, for cracks appearing within the substrate [34], due to their axial position differing from the silicon-air interface, their amplitude approaches zero when defocused. Consequently, when cracks occur in the silicon substrate, distinct imaging contrasts can be obtained based on the amplitude differences between the two signals [see Supplementary data 1].

## 3. Experimental setups and methods

### 3.1. Experimental setups

Fig. 1(a) presents a schematic representation of our in-line monitoring setup employing PARS technology. The comprehensive apparatus integrates both automation and the PARS system. The automation component features a robotic arm coupled with linear guides. Equipped with a visual positioning system, the robotic arm precisely identifies the sample placement positions, enabling efficient retrieval and placement actions. Upon receiving the feedback signal, the robotic arm checks the availability of vacant slots on the dual linear guide trays. If a slot is free,



**Fig. 1.** Setup and characterization of the PARS system for in-line defect monitoring. (a) Schematic of the in-line non-destructive defect monitoring scenario for flip-chip samples. (b) Schematic representation of the PARS system. (c) PARS image showing randomly distributed carbon fibers. (d) Characterization of lateral resolution, including line spread function (LSF) and edge spread function (ESF) derived from the data presented in panel (c). The resulting line spread functions depict the full width at half maximum (FWHM) lateral resolution.

the material tray is shifted to the left guide for loading retrieval. Once the tray is fully loaded, it transitions to the right, supplying the PARS system for continuous in-line defect monitoring experiments. The dual-guide design ensures seamless material supply. After the experiments, the qualified and defective products are sorted and placed into designated trays for subsequent processing.

As the pivotal components of the entire system, the PARS system primarily encompasses the excitation pathway, detection pathway, and data processing procedures, which closely resemble the system structure reported in our previous work [32]. In the excitation pathway, we employed an ultrashort pulsed laser (FSLAB-PICO-1064, Laize Photoelectric Technology, Beijing) operating at a wavelength of 1064 nm. This laser boasts an approximate pulse duration of 8 ps and a repetition frequency of 1 MHz, facilitating the generation of the excitation light beam. Initially, the laser beam was divided into two separate paths using a 9:1 beam splitter (BS as illustrated in Fig. 1(b)). In this configuration, 10 % of the optical power was directed towards a photodiode (PD-1 in Fig. 1(b), DET10A, Thorlabs) to generate trigger signals. Meanwhile, the transmitted light beam, comprising 90 % of the initial power, traversed through a beam expander (BE-1 in Fig. 1(b)) to attain the desired beam diameter of approximately 6 mm, and subsequently passed through a neutral density filter (NDF) for precise energy adjustment.

In the detection pathway, the probe beam was derived from a low-coherence 1310 nm continuous-wave source, utilizing superluminescent diodes (SLD1018P, Thorlabs) with a bandwidth of 45 nm. After expansion through a beam expander (BE-2 in Fig. 1(b)), the probe beam, with a diameter of 6 mm, propagated through a polarizing beam splitter (PBS, as depicted in Fig. 1(b), PBS254/M, Thorlabs) and a

quarter-wave plate (QWP, shown in Fig. 1(b), WPQ10M-1310, Thorlabs), resulting in the production of circularly polarized light. Following this, a dichroic mirror (DM, as illustrated in Fig. 1(b), DMSP1180, Thorlabs) was utilized to combine the probe beam with the excitation beam. Subsequently, both the probe and excitation beams were guided through the dual-axis galvanometer scanning mirror system (GM, illustrated in Fig. 1(b), GVS012/M, Thorlabs) and co-focused by the objective lens (OL, as presented in Fig. 1(b), MY20X-824, Mitutoyo).

In the data processing and digitalization stage, the back-reflected component of the probe beam retraced its path within the system and was directed to a photodiode (PD-2 in Fig. 1(b), 1811-FS, New Focus) for photoelectric conversion, subsequent to passing through a long-pass filter (LPF, as depicted in Fig. 1(b), FEL1250, Thorlabs) and a best-form lens (BFL, illustrated in Fig. 1(b), LBF254-040, Thorlabs). Following this, the photodiode's output was subjected to further filtration using a low-pass filter (22 MHz). The refined signal underwent digitalization through a four-channel, 14-bit PCI digitizer (CES1442, GaGe) operating at a rate of  $200 \times 10^6$  samples per second.

To evaluate the lateral resolution of the proposed system, we conducted cross-sectional scanning (B-scan) of carbon fiber networks in water using the PARS system. The imaging results are illustrated in Fig. 1(c). To determine the spatial resolution of the system, the line spread function (LSF) of the system was acquired. During the scanning process, particular attention was paid to the signal amplitude distribution at the edges of the carbon fibers. This distribution revealed the transition of the signal from absent to present within the B-scan. Fitting this distribution with an error Gaussian function provided the edge spread function (ESF) of the system. Differentiating the ESF yielded the Gaussian-shaped LSF.

The full width at half maximum (FWHM) value of the LSF in Fig. 1(d) is approximately  $2.9 \mu\text{m}$ , representing the lateral resolution of the designed system.

To achieve rapid and wide FOV imaging of flip chips during the fabrication process, we previously developed a wide-FOV imaging method based on an optical-mechanical joint scanning strategy in our earlier work [32]. Specifically, this method stitches together multiple smaller FOVs to create a larger FOV for expansive imaging. Each smaller FOV is scanned using a galvanometer scanning. Subsequently, motorized stages (M-112.2DG1, Physik Instrumente) move in specific steps to the adjacent smaller FOV for a similar galvanometer scanning, continuing this process until all smaller FOVs have been scanned. Finally, all the images from the individual smaller FOVs are stitched together to obtain a complete large FOV image of the sample [see Supplementary data 2].

### 3.2. PARS real-time imaging method based on continuous acquisition combined with parallel processing image reconstruction

In order to achieve the real-time imaging of the flip-chip sample by PARS technology, we proposed a real-time PARS imaging method based on continuous acquisition combined with GPU parallel processing image reconstruction. The continuous acquisition method is based on the data streaming mode of the data acquisition card. It continuously collects data at a sampling frequency of 200 MHz from the first trigger signal of the excitation light until predefined data volume is reached. The data utilization of this method is directly proportional to the repetition frequency of the excitation light. Therefore, a laser with a MHz repetition frequency can reduce the sampling time and enhance the imaging speed [see Supplementary data 3].

However, due to the continuous acquisition method resulting in larger data volume compared to existing trigger-based acquisition method [see Supplementary data 3], central processing unit (CPU)-based serial processing image reconstruction method is unable to meet the requirements of real-time imaging. Consequently, we proposed an image reconstruction method based on GPU parallel processing to reduce image reconstruction time and achieve real-time imaging with the PARS system [see Supplementary data 3].

Fig. 2(a) depicts the processing workflow of the CPU-based serial processing image reconstruction method under continuous acquisition. The method exhibits a time complexity of  $O(n)$ , as each step relies on the results from the previous step as input parameters, and the computations

within each step are limited to single or a few threads. Addressing this limitation, the GPU-based parallel processing image reconstruction method employs compute unified device architecture (CUDA) to parallelize computations at each step, distributing the workload across millions of threads. While this optimization method significantly enhances computational efficiency and achieves a time complexity of  $O(1)$ , the host-to-device memory transfer introduces additional memory copy time [35]. To overcome this, we implemented pinned memory technology and asynchronous GPU data transmission to ensure GPU memory copying occurs within the data transmission window of the acquisition card, thereby mitigating additional time consumption.

Furthermore, we present the strategy for solving the memory contention problem in parallel processing image reconstruction method by employing reduction algorithms and atomic operations, as shown in Fig. 2(b). The purpose of reduction algorithms is to find extreme values. The specific process involves dividing the data sent to the GPU memory into smaller data blocks, where each thread calculates the maximum and minimum values for its corresponding data block. Finally, these maximum and minimum values from all data blocks are consolidated on the CPU. Meanwhile, as the smallest operational unit, atomic operations of CUDA prohibit other parallel threads from performing read and write operations on the object variable during execution [36]. As a result, atomic operations effectively block threads that encounter memory contention, thereby achieving mutual exclusion protection for shared variables among threads and becoming the safest method for resolving memory contention, as shown in Fig. 2(b).

However, when a protected shared variable needs to be accessed by each thread, each thread becomes blocked, resulting in parallel threads executing serially, thus leading to significant time consumption. Based on this mechanism, for efficiently protecting various shared variables, we developed a multi-atomic operation algorithm by using an array as the object of atomic operations. In this scenario, each element in the array serves solely as a lock for managing thread blocking and execution states, without participating in computation. In the implementation of the algorithm, the number of elements in the array is equal to the number of pixels in the PARS image. When memory contention occurs in the computation of a pixel, only the lock corresponding to that pixel relies on the atomic operation to block threads in competition, while parallel threads computing other pixels remain unblocked, as shown in Fig. 2(b).

We utilized CUDA (v11.3) for programming and implemented the GPU-based parallel processing image reconstruction method on an i7-

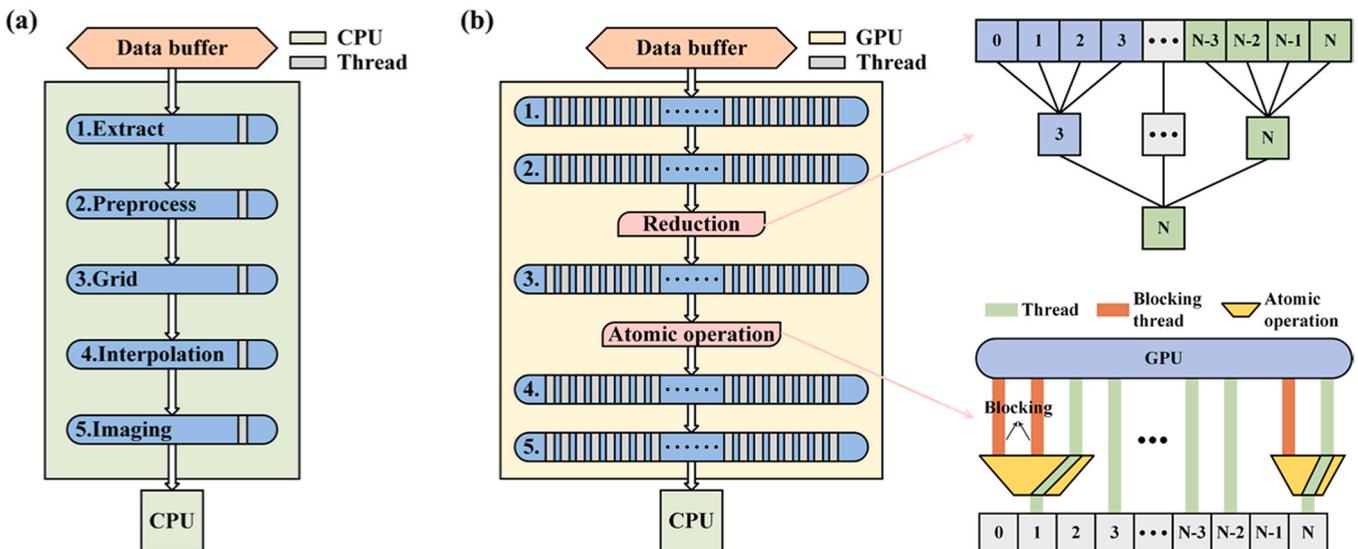


Fig. 2. The process of PARS image reconstruction. (a) The CPU-based serial processing image reconstruction method. (b) The GPU-based parallel processing image reconstruction method.

8700 CPU and a single NVIDIA GeForce RTX 2080Ti GPU. Furthermore, this method, as the real-time imaging method for the PARS system, has been integrated into the PARS Video Mode Studio software, which is used for system control and configuration. This software is programmed in QT (v5.14.2) with C++ and incorporates features such as the user interface, imaging, motorized-stage control, scanning-mirror control, file saving, and more.

### 3.3. Precision in-line recognition method for microscopic defects based on YOLOv7-SPD

To enhance the in-line recognition accuracy of microscopic defects in the flip-chip sample by PARS technology, we developed an IYOLOv7-SPD convolutional neural network based on YOLOv7, and introduced a convolutional neural network (CNN) building block called space to depth (SPD). The SPD block effectively addresses the common issue in CNN architectures, where strided convolutions or pooling layers lead to the loss of fine-grained information and learning of less effective feature representations. This enhancement enables the network to more effectively extract and preserve feature information related to microscopic defects, thereby boosting its capability for microscopic defect recognition.

The building block consists of a space-to-depth layer and a non-strided convolution layer. Consider any input feature map  $X$  of size  $S \times S \times C$ , downsampling is performed using a factor of scale. Fig. 3(a) provides an example when  $scale = 2$ , where we obtain four sub-maps each of which is of size  $S/2, S/2, C$  and downsamples  $X$  by a factor of 2. These sub-maps are then concatenated along the channel dimension to obtain a feature map  $X'$ . The spatial dimensions of  $X'$  are reduced by a factor of scale, while the channel dimension is increased by a factor of  $scale^2$ , preserving all the information along the channel dimension without loss. Following the space-to-depth feature transformation layer, a non-strided (stride = 1) convolution layer with  $C'$  filters is added to transform the feature map  $X'$  ( $S/scale, S/scale, scale^2C$ ) into the feature map  $X''$  ( $S/scale, S/scale, C'$ ). The non-strided convolution aims to retain all discriminative feature information as much as possible.

Fig. 3(b) illustrates the structure of the IYOLOv7-SPD defect recognition method. We obtained IYOLOv7-SPD by replacing the strided-2 convolution layers with SPD block in YOLOv7. It retains the overall network architecture of YOLOv7 and is a typical one-stage object detection model [37]. The model consists of a backbone for feature extraction and a detection head for predicting the class and bounding box of each object. In between, a neck of extra layers is added to combine features at multiple scales to produce semantically strong features for detecting objects of different sizes. The spatial pyramid pooling cross stage partial connection (SPPCSPC) block in the backbone incorporates four different-scale max-pooling layers, enhancing the

receptive field to adapt to objects of varying scales. The concat block concatenates outputs from different layers, the upsample (Up) block performs upsampling using nearest-neighbor interpolation, the max pooling (MP) block achieves downsampling through convolution and max-pooling layers, and the detect block contains different-sized detection heads.

Additionally, we implemented our network using the Pytorch (v1.12.0) library and conducted all experiments on a single NVIDIA GeForce RTX 2080Ti. On the basis of the training set, we determined our bounding box priors using k-means++ clustering [38]. The Adam optimizer [39] was used to optimize our network with an initial learning rate of 0.0001, a batch size of 2 and 200 epochs. The loss function of our network consisted of classification loss, confidence loss, and bounding box loss. Binary cross-entropy loss was used to calculate classification loss and confidence loss. CIoU loss [40] was adopted to calculate bounding box loss.

### 3.4. Sample preparation

To demonstrate the in-line non-destructive defect monitoring capabilities of our PARS system, a flip-chip sample was fabricated by depositing a 200 nm-thick layer of aluminum onto a silicon substrate. The sample preparation processes involved spin coating, dehydration, photolithography, development, evaporation, and lift-off, as shown in Fig. S4. To simulate internal defects within the flip-chip, such as delamination [33,34], we employed ultrasonic cleaning machine (SK2200LHC, Shanghai KeDao Ultrasonic Instrument Co., Ltd.) to perform ultrasonic cleaning on the prepared flip-chip samples to induce partial peeling of the metallic structures, as shown in Fig. S6. The ultrasonic frequency and energy used were 53 kHz and 90 W, respectively. Additionally, cracks were introduced in the silicon regions without metal coating to mimic possible cracks in the flip-chip substrate [34], as shown in Fig. S6. Throughout the entire experimental process, the flip chip sample was reversed and placed on the motorized stages, ensuring that the metal structures and the cracks were consistently positioned beneath the silicon wafer, as illustrated in the insert of Fig. 1(b). This implies that all defects were invisible under brightfield microscopy. The experimental process simulated the scenario of in-line monitoring internal defects within flip chip samples.

## 4. Results

### 4.1. Real-time imaging of the flip-chip sample based on PARS technology

To demonstrate the reliability and real-time imaging capability of the proposed imaging method of PARS, we conducted optical-mechanical joint scanning imaging experiments on the flip-chip

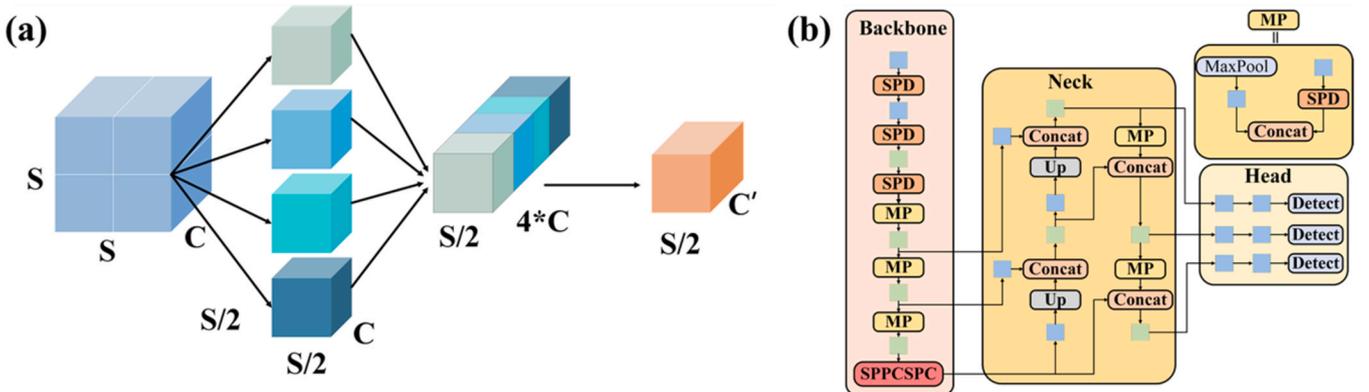


Fig. 3. Overview of IYOLOv7-SPD. (a) Illustration of SPD block when  $scale = 2$ . (b) Detailed structure of IYOLOv7-SPD. Blue blocks represent convolution layers. Green blocks represent Extended efficient layer aggregation networks.

sample. During the experiment, the data acquisition card operated in data streaming mode and continuously collected data [see [Supplementary data 3](#)] at a sampling frequency of 200 MHz, with a configured sampling depth of  $8 \times 10^7$ . In the optical-mechanical joint scanning imaging mode, the scanning frequencies of the dual-axis galvanometer mirrors were set at 2.5 Hz and 800 Hz, respectively, and remained constant throughout all experiments, yielding a small FOV of  $400 \times 400 \mu\text{m}^2$ . The motorized stages scanned sequentially in the x and y directions at a speed of 1.5 mm/s, with a single-step movement of  $320 \mu\text{m}$  [see [Supplementary data 2](#)]. This resulted in a total of  $4 \times 2$  small FOVs from the galvanometer scanning, with the entire process completed in 8 seconds.

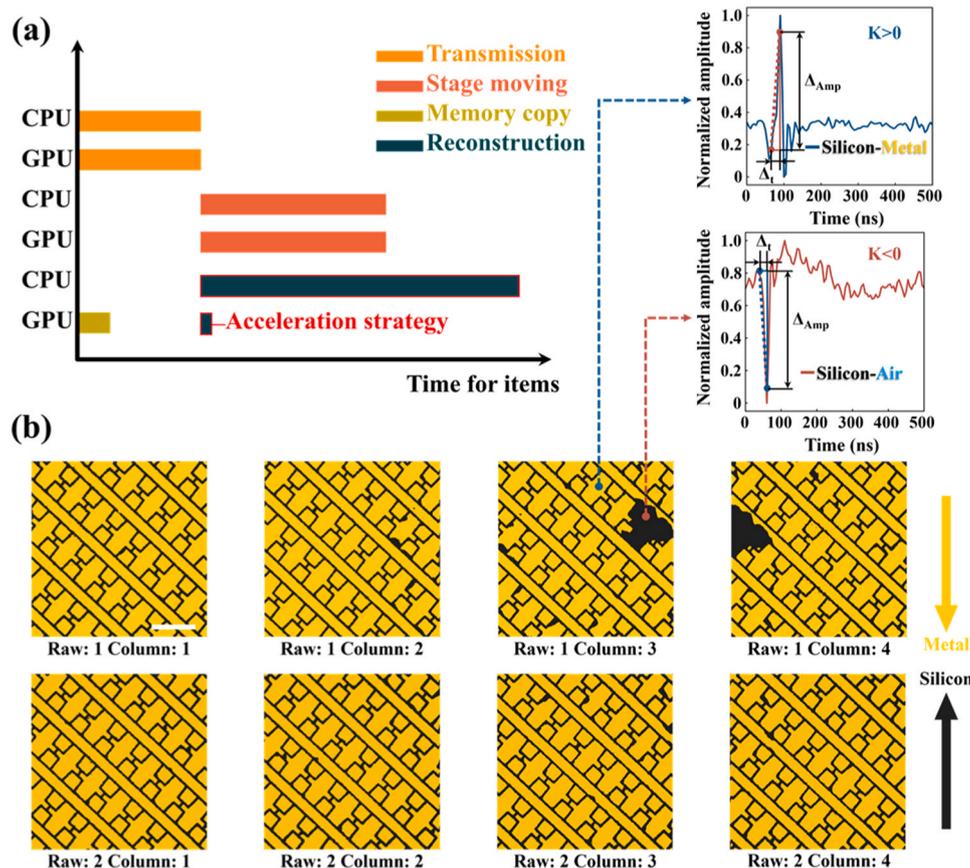
[Fig. 4\(a\)](#) compares the time consumption of each workflow when the PARS system uses the continuous acquisition combined with CPU-based serial processing image reconstruction method and the PARS real-time imaging method. In the PARS real-time imaging method, the overall workflow includes the transmission phase, motorized stages moving phase, memory copy phase, and reconstruction phase. After the PARS system is launched, it enters the transmission phase. In this phase, the time consumption for both methods is 400 ms, as the transmission time is determined by the set sampling quantity of the acquisition card and is independent of the method used. Subsequently, during the movement of the motorized stages, the PARS system completes image reconstruction. The time for the motorized stages movement is not affected by the method and is approximately 600 ms.

In the image reconstruction phase, the average consumption time for the CPU-based serial processing image reconstruction method is 1134 ms, which is greater than the time for the motorized stages movement. In contrast, the average consumption time for the GPU-based parallel processing image reconstruction method is only 38 ms, which is 30 times shorter than the CPU-based serial processing image

reconstruction method. This ensures that image reconstruction in a GPU environment can be completed during the movement of the motorized stage. The memory copy phase is inevitable when using the GPU-based parallel processing image reconstruction method. However, thanks to pinned memory technology and asynchronous data transmission technology, the memory copy phase only takes 125 ms and can run concurrently with the transmission phase. Therefore, the memory copy phase does not impact the imaging time of the PARS system.

[Fig. 4\(b\)](#) displays the optical-mechanical joint scanning results (2 rows  $\times$  4 columns) of the flip-chip sample obtained using the PARS real-time imaging method, which is based on continuous acquisition combined with parallel processing image reconstruction. While the imaging results are consistent with those obtained using the continuous acquisition combined with CPU-based serial processing image reconstruction method [see [Supplementary data 5](#)], the proposed method demonstrates an average reconstruction time that is approximately 30 times faster. The reconstruction times of both methods during optical-mechanical joint scanning are summarized in [Table 1](#).

This speed improvement allows the optical-mechanical joint scanning mode to no longer be limited by the image reconstruction time, thereby contributing to the achievement of real-time imaging for flip-chip samples. Due to the smaller step size of the motorized stages during optical-mechanical joint scanning than the FOV of the image, adjacent images have a 20 % overlap region. The imaging results are the same in the overlap regions of adjacent images, which confirms the continuity and reliability of PARS real-time imaging method. The experimental results accurately display defects of different sizes, regions, and shapes, confirming the accuracy of real-time imaging in the PARS system. Additionally, Video1 in the Supplementary data demonstrates the real-time imaging process of this experiment. The results indicate that the PARS real-time imaging method based on continuous



**Fig. 4.** The real-time imaging results for the flip-chip sample. (a) Timeline chart of time consumption for various components in the PARS system (b) PARS optical-mechanical joint scanning imaging results based on the proposed real-time imaging method. All images are presented at the same scale bar: 100  $\mu\text{m}$ .

**Table 1**

Comparison of image reconstruction times based on two imaging methods.

Image coordinate (Raw, Column)		(1, 1)	(1, 2)	(1, 3)	(1,4)	(2, 1)	(2, 2)	(2, 3)	(2, 4)
Reconstruction time (ms)	Continuous acquisition combined with CPU-based serial processing	1140	1136	1148	1127	1134	1130	1125	1131
	PARS real-time imaging	35	37	43	36	36	38	43	37

acquisition combined with parallel processing image reconstruction is promising for the real-time imaging of flip-chip structures during the fabrication process.

Supplementary material related to this article can be found online at [doi:10.1016/j.pacs.2024.100614](https://doi.org/10.1016/j.pacs.2024.100614).

#### 4.2. In-line accurate defect recognition of flip chip samples based on IYOLOV7-SPD

To validate the accurate and fast defect recognition capabilities of the proposed YOLOv7-SPD, we established the dataset of defects in flip-chip samples. Specifically, we captured 4075 defect images using the constructed PARS system. Subsequently, we manually annotated the defect categories and positional information in the images using the Labelling tool, thus creating the defect dataset. According to the object definition method based on absolute scale [41–43], the dataset comprises 27,726 microscopic defects, 7386 medium defects, and 1078 large defects. Here, the term "microscopic defects" is defined as delamination defects with a pixel area smaller than  $32 \times 32 \text{ pixel}^2$ , corresponding to an actual area of less than  $48 \times 48 \mu\text{m}^2$  when converted [see Supplementary data 6]. To alleviate the overfitting problem and improve the generalization ability of our network, we employed data augmentation techniques such as affine transformation, horizontal flipping, color jittering and pixel multiplication [44]. The augmented dataset included 20,375 images and was randomly split into a train set and a test set in a 9:1 ratio.

To validate the superior recognition capability of our proposed IYOLOv7-SPD for detecting microscopic defects, we compared it with classical deep learning-based object detection methods including Faster RCNN [45], YOLOv3 [46], YOLOX [47], and YOLOv7 [48]. We trained these methods on the same training dataset and computed their recognition accuracy on an identical test dataset, as shown in Table 2.

Table 2 introduces the mean average precision (mAP) as a comprehensive metric to assess the accuracy of the recognition methods [49]. Through this metric, we compared the recognition accuracy of IYOLOv7-SPD with other methods specifically on microscopic defects of delamination. For microscopic defects in the flip-chip samples, IYOLOv7-SPD achieved mAP of 56.8 %, which is the highest recognition accuracy among the five methods compared, demonstrating the superior capability of IYOLOv7-SPD in recognizing microscopic defects accurately. Moreover, with a single recognition time of 12 ms, IYOLOv7-SPD ensures precise in-line recognition of microscopic defects during real-time imaging.

Furthermore, to demonstrate the in-line accurate defect recognition capability of IYOLOv7-SPD for microscopic defects more intuitively, we employed various recognition methods to identify defects from the large-scale scanning image of the flip-chip sample. The image was obtained by setting the scanning range of the galvanometer mirrors to  $600 \times 600 \mu\text{m}^2$  with a scanning step size of  $3.2 \mu\text{m}$ . Fig. 5(b–e,g) compare the microscopic defect recognition results obtained by

**Table 2**

Comparison of recognition accuracy of microscopic defects by different recognition methods.

Method	Faster RCNN	YOLOv3	YOLOX	YOLOv7	IYOLOv7-SPD
mAP	54.3 %	48.7 %	52.5 %	54.6 %	56.8 %
Time	42 ms	28 ms	16 ms	12 ms	12 ms

IYOLOv7-SPD with those of other recognition methods using the large-scale scanning image. The recognition results confirm that the confidence level of IYOLOv7-SPD in recognizing microscopic defects is higher than that of other recognition methods and can recognize microscopic defects missed by other methods. Furthermore, compared to the signal gating methods, the proposed method can effectively differentiate between delamination and substrate structures in Fig. 5(f).

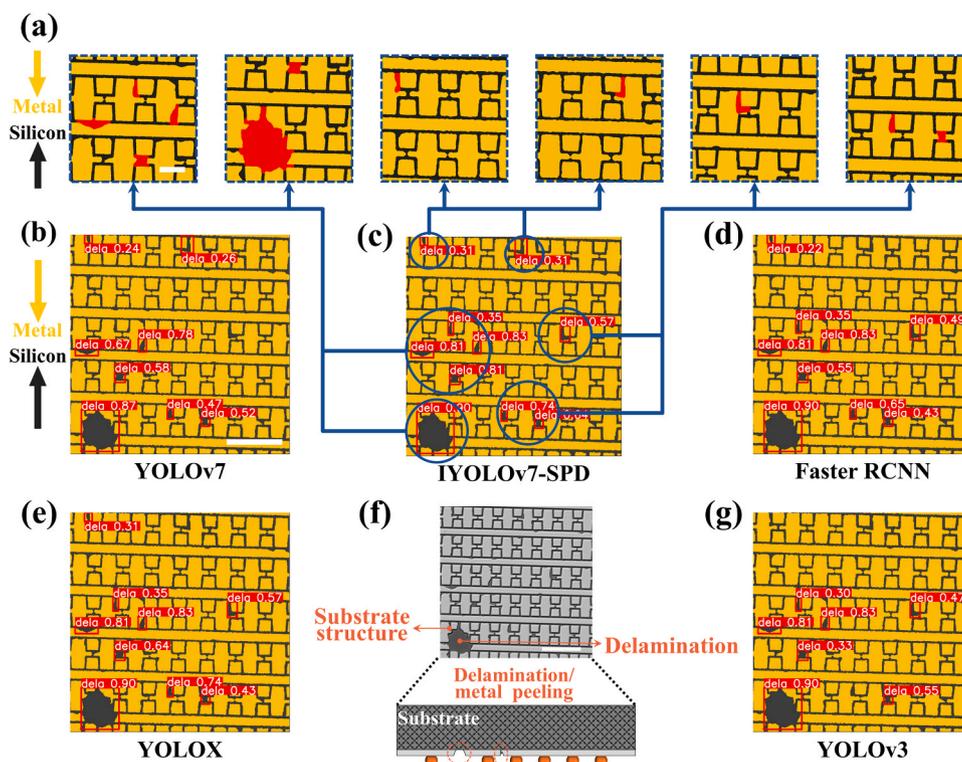
To validate the reliability of the results achieved by IYOLOv7-SPD in large-scale recognition, we performed a small-scale scanning using its recognition outcomes. This small-scale scanning features a finer step size and enhanced lateral resolution [see Supplementary data 9]. Specifically, by controlling motorized stages to move to the locations of different defects identified by IYOLOv7-SPD, we performed small-scale scanning over an area of  $240 \times 240 \mu\text{m}^2$  with a scanning step size of  $1.3 \mu\text{m}$  and manually segmented the regions with defects in the images, as shown in Fig. 5(a). It can be observed that the defect imaging results from the small-scale scanning match those of the large-scale recognition. Furthermore, upon completion of the experiment, we acquired bright-field microscopy image of the corresponding sample regions for further validation, as depicted in Fig. 5(f).

Based on the aforementioned results, it is evident that IYOLOv7-SPD possesses the capability to recognize microscopic defects in flip-chip samples and demonstrates a higher precision in microscopic defect recognition compared to other methods. Additionally, this method also demonstrates higher recognition accuracy for medium and large-sized defects compared to other deep learning approaches [see Supplementary data 7]. On the other hand, beyond the delamination in the chip samples, our proposed system also exhibits the capability for in-line non-destructive defect monitoring of substrate cracks in flip-chip samples [see Supplementary data 8]. These findings underscore the potential and versatility of our proposed PARS in-line defect monitoring system in enhancing quality control and defect monitoring processes in flip-chip manufacturing.

## 5. Discussion and conclusion

In this study, we proposed a PARS-based in-line non-destructive defect monitoring system designed for real-time imaging of flip chips during the fabrication process and for accurate defect recognition of their internal defects. To meet the requirements of real-time imaging and accurate recognition of microscopic defects for in-line non-destructive defect monitoring, we proposed real-time imaging and microscopic defect recognition methods. These methods were applied to achieve real-time imaging of the fabricated flip-chip samples and in-line precise recognition of microscopic defects. The results demonstrate that our proposed PARS system has the potential for in-line non-destructive and precise defect monitoring of flip-chips during the fabrication process, showcasing its unique advantages compared to other techniques, as presented in Table 3.

To achieve real-time imaging, we proposed real-time imaging method of PARS based on continuous acquisition combined with parallel processing image reconstruction that ensures the imaging process does not increase the optical-mechanical scanning time, while maintaining the same micrometer-level lateral resolution as the method based on continuous acquisition combined with serial processing. This method reduces the image reconstruction time from an average of approximately 1134 ms to around 38 ms, achieving a 30-fold improvement compared to the CPU-based serial processing image reconstruction method under



**Fig. 5.** Comparative experimental results of different recognition methods. (a) Small-scale precise scanning result. (b) Recognition result of YOLOv7. (c) Recognition result of IYOLOv7-SPD. (d) Recognition result of Faster RCNN. (e) Recognition result of YOLOX. (f) Brightfield microscopy imaging result. (g) Recognition result of YOLOv3. The recognition results are denoted by red rectangular boxes. The regions of small-scale precise scanning are highlighted with blue circular boxes. Defect areas within the small-scale precise scanning results are delineated in red. Scale bar for large-scale and brightfield images: 150  $\mu\text{m}$ . Scale bar for small-scale images: 50  $\mu\text{m}$ .

continuous acquisition. The compatibility of PARS technology’s real-time imaging capability with the optical-mechanical joint scanning mode makes it particularly well-suited for the in-line imaging of flip chips during the fabrication process.

To achieve precise recognition of microscopic defects within flip-chip samples during real-time imaging, we introduce the IYOLOv7-SPD recognition approach. This method incorporates SPD block, to mitigate the use of strided convolution or pooling layers, which results in a loss of fine-grained information and learning of less effective feature representations. Consequently, compared to other deep learning-based methods, IYOLOv7-SPD enables more accurate in-line recognition of microscopic defects.

In the future, to achieve faster imaging, higher-speed motorized stages can be employed in conjunction with galvo-resonant scanners. Concurrently, the development of corresponding image compensation algorithms may be necessary to ensure that the signal-to-noise ratio of the images remains unaffected by mechanical instabilities during rapid scanning [see [Supplemental data 10](#)]. To achieve higher precision in identifying microscopic defects, the neural network structure of IYOLOv7-SPD can be further optimized by integrating a context enhancement module. Additionally, enhancing the spatial resolution of PARS can be achieved by incorporating a confocal optical design in the detection path. This would increase the pixel area occupied by microscopic defects in the images, thereby reducing the difficulty in their accurate identification. On the other hand, the combination of PARS with laser ultrasound techniques [22–26] holds promise for non-destructive monitoring of various defect types in flip chips across multiple scales, such as voids and delamination between the die top and the encapsulates.

**CRedit authorship contribution statement**

**KAIXUAN DING:** Software, Investigation, Formal analysis. **YIHAN PI:** Writing – original draft, Visualization, Validation. **JIJING CHEN:** Writing – original draft, Validation, Methodology, Data curation. **ZHEN TIAN:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **SHOUJUN ZHANG:** Resources. **jiao li:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

**Appendix A. Supporting information**

Supplementary data associated with this article demonstrates the real-time imaging of the flip-chip sample using the PARS technology in conjunction with the PARS real-time imaging method and other experimental details.

**Funding**

This work was supported by National Natural Science Foundation of China (82171989, 62235013); Tianjin Municipal Fund for Distinguished Young Scholars (20JCJQJC00190); International Science and Technology Independent Cooperation Project of Shenzhen (GJHZ20210705142401004); Key Fund of Shenzhen Natural Science Foundation (JCYJ20200109150212515); 2022 Shenzhen College and University Stability Support Plan.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Table 3**  
Summary of non-destructive testing techniques for flip chips.

Methods	Advantages	Disadvantages	In-line	Microscopic defect recognition
Ultrasonic [8–11]	Nondestructive Deep penetration depth	Require liquid coupling Micron resolution not achieved	No	No
SAW [12, 13]	Nondestructive	Trade-off between resolution and imaging depth	No	No
X-CT [14–16]	Noncontact Sub-micron 3D imaging	Requires time-consuming data processing Ionizing radiation	No	Yes
PPT [17,18]	Noncontact Nondestructive	Complex and time-consuming data processing procedure Micron resolution not achieved	Yes	No
Traditional PA [19–21]	Micron resolution Deep penetration depth	Require liquid coupling	No	Yes
LU [22–26]	Deep penetration depth High SNR Noncontact Nondestructive	Micron resolution not achieved	Yes	No
Previous PARS [27–32]	Micron resolution Deep penetration depth Noncontact Nondestructive	Long image reconstruction time Unable to intelligently recognize	No	Yes
Our PARS	Deep penetration depth Noncontact Nondestructive Fast imaging speed Intelligent recognition	Currently limited by the scanning speed of the motorized stages	Yes	Yes

**Data availability**

Data will be made available on request.

**Appendix A. Supporting information**

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.pacs.2024.100614.

**References**

[1] P.L. Martin. *Electronic Failure Analysis Handbook: Techniques and Applications for Electronic and Electrical Packages, Components, and Assemblies*, 1st edition, McGraw-Hill Education, New York, 1999. (<https://www.accessengineeringlibrary.com/content/book/9780070410442>).

[2] K.C. Cheng, K.S. Li, A.Y. Huang, J.W. Li, L.L. Chen, N.C. Tsai, S.J. Wang, C.S. Lee, L. Chou, P.Y.-Y. Liao, H.C. Liang, J.E. Chen, Wafer-level test path pattern recognition and test characteristics for test-induced defect diagnosis, in: *Proceedings of the 2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, IEEE, 2020, pp. 1710–1. (<https://doi.org/10.23919/date48585.2020.9116546>).

[3] B. Kumar, K. Basu, M. Fujita, V. Singh, Post-silicon gate-level error localization with effective and combined trace signal selection, *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* 39 (2018) 248–261, <https://doi.org/10.1109/tcad.2018.2883899>.

[4] K.J.P. Jacobs, Y. Li, M. Stucchi, I.D. Wolf, S.V. Huylenbroeck, J.D. Vos, E. Beyne, Optical beam-based defect localization methodologies for open and short failures

in micrometer-scale 3-D TSV interconnects, *IEEE Trans. Compon. Hybrids Manuf. Technol.* 10 (2020) 1542–1551, <https://doi.org/10.1109/TCPMT.2020.3014161>.

[5] S. Seal, M.D. Glover, H.A. Mantooh, 3-D wire bondless switching cell using flip-chip-bonded silicon carbide power devices, *IEEE Trans. Power Electron.* 33 (2017) 8553–8564, <https://doi.org/10.1109/tpel.2017.2782226>.

[6] X. Gu, D. Liu, C. Baks, O. Tageman, B. Sadhu, J. Hallin, L. Rexberg, P. Parida, Y. Kwarik, A. Valdes-Garcia, Development, implementation, and characterization of a 64-element dual-polarized phased-array antenna module for 28-GHz high-speed data communications, *IEEE Trans. Microw. Theory Tech.* 67 (2019) 2975–2984, <https://doi.org/10.1109/tmtt.2019.2912819>.

[7] T. Matsumoto, T. Kurahashi, R. Konoike, K. Suzuki, K. Tanizawa, A. Uetake, K. Takabayashi, K. Ikeda, H. Kawashima, S. Akiyama, S. Sekiguchi, Hybrid-integration of SOA on silicon photonics platform based on flip-chip bonding, *J. Lightw. Technol.* 37 (2018) 307–313, <https://doi.org/10.1109/jlt.2018.2870128>.

[8] L. Su, S. Tan, Y. Qi, J. Gu, Y. Ji, G. Wang, X. Ming, K. Li, M. Pecht, An improved orthogonal matching pursuit method for denoising high-frequency ultrasonic detection signals of flip chips, *Mech. Syst. Signal. Pract.* 188 (2023) 110030, <https://doi.org/10.1016/j.ymsp.2022.110030>.

[9] C. Lee, G. Zhang, D.M. Harvey, A. Qi, Characterization of micro-crack propagation through analysis of edge effect in acoustic microimaging of microelectronic packages, *NDTE Int.* 79 (2016) 1–6, <https://doi.org/10.1016/j.ndteint.2015.11.007>.

[10] B. Wu, Y. Huang, S. Krishnaswamy, A Bayesian approach for sparse flaw detection from noisy signals for ultrasonic NDT, *NDTE Int.* 85 (2017) 76–85, <https://doi.org/10.1016/j.ndteint.2016.10.005>.

[11] L. Su, X. Yu, K. Li, X. Yao, M. Pecht, Simulation and experimental verification of edge blurring phenomenon in microdefect inspection based on high-frequency ultrasound, *IEEE Access* 7 (2019) 11515–11525, <https://doi.org/10.1109/access.2019.2892231>.

[12] Z. Liu, L. Bin, X. Liang, A. Du, Quantifying the subsurface damage and residual stress in ground silicon wafer using laser ultrasonic technology: a Bayesian approach, *Mech. Syst. Signal. Process.* 173 (2022) 109008, <https://doi.org/10.1016/j.ymsp.2022.109008>.

[13] Z. Liu, L. Bin, X. Liang, X. Ma, Y. Wan, Multimode photoacoustic characterization of subsurface damage in ground thin wafers, *Int. J. Mech. Sci.* 238 (2023) 107845, <https://doi.org/10.1016/j.ijmecsci.2022.107845>.

[14] L. Su, L. Wang, K. Li, J. Wu, G. Liao, T. Shi, T. Lin, Automated X-ray recognition of solder bump defects based on ensemble-ELM, *Sci. China: Technol. Sci.* 62 (2019) 1512–1519, <https://doi.org/10.1007/s11431-018-9324-3>.

[15] S. Imashuku, K. Wagatsuma, Non-destructive evaluation of alumina scale on heat-resistant steels using cathodoluminescence and X-ray-excited optical luminescence, *Corros. Sci.* 154 (2019) 226–230, <https://doi.org/10.1016/j.corsci.2019.04.009>.

[16] L. Yi, B. Hou, H. Zhao, H. Tan, X. Liu, A double-tapered fibre array for pixel-dense gamma-ray imaging, *Nat. Photonics* 17 (2023) 494, <https://doi.org/10.1038/s41566-023-01204-1>.

[17] X. Lu, G. Liao, Z. Zha, Q. Xia, T. Shi, A novel approach for flip chip solder joint inspection based on pulsed phase thermography, *NDTE Int.* 44 (2011) 484–489, <https://doi.org/10.1016/j.ndteint.2011.05.003>.

[18] L. Alam, N. Kehtarnavaz, A survey of detection methods for die attachment and wire bonding defects in integrated circuit manufacturing, *IEEE Access* 10 (2022) 83826–83840, <https://doi.org/10.1109/access.2022.3197624>.

[19] S. Jeon, J. Kim, J.P. Yun, C. Kim, Non-destructive photoacoustic imaging of metal surface defects, *J. Opt.* 18 (2016) 114001, <https://doi.org/10.1088/2040-8978/18/11/114001>.

[20] H. Liu, Y. Zhao, J. Zhou, P. Li, S. Bo, S. Chen, Photoacoustic imaging of lithium metal batteries, *ACS Appl. Energy Mater.* 3 (2019) 1260–1264, <https://doi.org/10.1021/acsaem.9b01791>.

[21] M. Sun, X. Cheng, Y. Wang, X. Zhang, Y. Shen, N. Feng, Method for detecting high-speed rail surface defects by photoacoustic signal, *Acta Phys. Sin.* 65 (2016) 038105, <https://doi.org/10.7498/aps.65.038105>.

[22] I. Pelivanov, L. Ambrozinski, A. Khomenko, E.G. Koricho, G.L. Cloud, M. Haq, M. O'Donnell, High resolution imaging of impacted CFRP composites with a fiber-optic laser-ultrasound scanner, *Photoacoustics* 4 (2016) 55–64, <https://doi.org/10.1016/j.pacs.2016.05.002>.

[23] I. Pelivanov, M. O'Donnell, Imaging of porosity in fiber-reinforced composites with a fiber-optic pump-probe laser-ultrasound system, *Compos. Part A Appl. Sci. Manuf.* 79 (2015) 43–51, <https://doi.org/10.1016/j.compositesa.2015.09.014>.

[24] P. Pyzik, A. Ziaja-Sujdak, J. Spytek, M. O'Donnell, I. Pelivanov, L. Ambrozinski, Detection of disbonds in adhesively bonded aluminum plates using laser-generated shear acoustic waves, *Photoacoustics* 21 (2021) 100226, <https://doi.org/10.1016/j.pacs.2020.100226>.

[25] V.V.B. Reddy, I.C. Ume, A. Mebane, K. Akinade, B. Rogers, C. Guirguis, P.B. Patel, K. Derksen, R. Case, Evaluation of FCBGa package subjected to four-point bend reliability test using fiber array laser ultrasonic inspection system, *IEEE Trans. Compon. Hybrids Manuf. Technol.* 9 (2019) 1219–1226, <https://doi.org/10.1109/TCPMT.2019.2911073>.

[26] V.V.B. Reddy, I.C. Ume, J. Williamson, S.K. Sitaraman, Evaluation of the quality of BGA solder balls in FCBGa packages subjected to thermal cycling reliability test using laser ultrasonic inspection technique, *IEEE Trans. Compon. Hybrids Manuf. Technol.* 11 (2011) 589–597, <https://doi.org/10.1109/TCPMT.2021.3065958>.

[27] P. Hajireza, W. Shi, K. Bell, R.J. Paproski, R.J. Zemp, Non-interferometric photoacoustic remote sensing microscopy, *Lightw.: Sci. Appl.* 6 (2017) e16278, <https://doi.org/10.1038/lsa.2016.278>.

[28] L. Mukhangaliyeva, S. Kocer, A. Warren, K. Bell, M. Boktor, M. Yavuz, E. Abdel-Rahman, P.H. Reza, Deformable mirror-based photoacoustic remote sensing

- (PARS) microscopy for depth scanning, *Biomed. Opt. Express* 13 (2022) 5643–5653, <https://doi.org/10.1364/boe.471770>.
- [29] S. Abbasi, K. Bell, B. Ecclestone, P.H. Reza, Live feedback and 3D photoacoustic remote sensing, *Quant. Imaging Med. Surg.* 11 (2021) 1033, <https://doi.org/10.21037/qims-20-758>.
- [30] K.L. Bell, P.H. Reza, R.J. Zemp, Real-time functional photoacoustic remote sensing microscopy, *Opt. Lett.* 44 (2019) 3466–3469, <https://doi.org/10.1364/ol.44.003466>.
- [31] B.S. Restall, B.D. Cikaluk, M.T. Martell, N.J. Haven, R. Mittal, S. Silverman, L. Peiris, J. Deschenes, B.A. Adam, A. Kinnaird, R.J. Zemp, Fast hybrid optomechanical scanning photoacoustic remote sensing microscopy for virtual histology, *Biomed. Opt. Express* 13 (2022) 39–47, <https://doi.org/10.1364/boe.443751>.
- [32] J. Chen, S. Li, Y. Long, X. Chen, B. Liu, M. Hu, J. Li, Z. Tian, Nondestructive inspection of metallic microstructure chips based on photoacoustic remote sensing microscopy, *Appl. Phys. Lett.* 120 (2022) 182201, <https://doi.org/10.1063/5.0090897>.
- [33] C.M. Tan, E. Er, Y. Hua, V. Chai, Failure analysis of bond pad metal peeling using FIB and AFM, *IEEE Trans. Compon. Packag. Manuf. Technol. Part A* 21 (1998) 585–591, <https://doi.org/10.1109/95.740051>.
- [34] S.U. Fassbender, K. Kraemer, Acoustic microscopy – a powerful tool to inspect microstructures of electronic devices, *Test. Reliab. Appl. Micro- Nano-Mater. Syst.* 5045 (2003) 112–121, <https://doi.org/10.1117/12.483797>.
- [35] R.S. Santos, D.M. Eler, R.E. Garcia, Performance evaluation of data migration methods between the host and the device in CUDA-based programming, in: *Proceedings of the Information Technology: New Generations: 13th International Conference on Information Technology*, Springer, 2016, pp. 689–700. ([https://doi.org/10.1007/978-3-319-32467-8\\_60](https://doi.org/10.1007/978-3-319-32467-8_60)).
- [36] H.V. Dang, B. Schmidt, CUDA-enabled Sparse Matrix–Vector Multiplication on GPUs using atomic operations, *Parallel Comput.* 39 (2013) 737–750, <https://doi.org/10.1016/j.parco.2013.09.005>.
- [37] K. Li, X. Wang, H. Lin, L. Li, Y. Yang, C. Meng, J. Gao, Survey of one-stage small object detection methods in deep learning, *J. Front. Comput. Sci. Technol.* 16 (2022) 41. (<http://fcst.ceaj.org/CN/10.3778/j.issn.1673-9418.2110003>).
- [38] D. Arthur, S. Vassilvitskii, K-means++ the advantages of careful seeding, in: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007, pp. 1027–35. (<https://dl.acm.org/doi/10.5555/1283383.1283494>).
- [39] D.P. Kingma, J. Ba, Adam, A method for stochastic optimization, *arXiv Prepr. arXiv*, 2014, 1412.6980. (<https://arxiv.org/abs/1412.6980>).
- [40] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren, Distance-IoU loss: faster and better learning for bounding box regression, *Proc. AAAI Conf. Artif. Intell.* (2020) 12993–13000, <https://doi.org/10.1609/aaai.v34i07.6999>.
- [41] T.Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft COCO: common objects in context, *Lect. Notes Comput. Sci.* 8693 (2014) 740–755, [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48).
- [42] C. Yang, Z. Huang, N. Wang, QueryDet: cascaded sparse query for accelerating high-resolution small object detection, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 13658–67. (<https://doi.org/10.1109/CVPR52688.2022.01330>).
- [43] W. Wang, J. Dai, Z. Chen, Z. Huang, Z. Li, X. Zhu, X. Hu, T. Lu, L. Lu, H. Li, X. Wang, Y. Qiao, InternImage: exploring large-scale vision foundation models with deformable convolutions, in: *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 14408–19. (<https://doi.org/10.1109/CVPR52729.2023.01385>).
- [44] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, *J. Big Data* 6 (2019) 1–48, <https://doi.org/10.1186/s40537-019-0197-0>.
- [45] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (2017) 1137–1149, <https://doi.org/10.1109/TPAMI.2016.2577031>.
- [46] J. Redmon, A. Farhadi, YOLOv3: an Incremental Improvement, *arXiv Prepr. arXiv*, 2018, 1804.02767. (<https://doi.org/10.48550/arXiv.1804.02767>).
- [47] Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun, YOLOX: exceeding YOLO series in 2021, *arXiv Prepr. arXiv*, 2021, 2107.08430. (<https://doi.org/10.48550/arXiv.2107.08430>).
- [48] C.Y. Wang, A. Bochkovskiy, H.Y.M. Liao, YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in: *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 7464–75. (<https://doi.org/10.1109/CVPR52729.2023.00721>).
- [49] G. Chen, H. Wang, K. Chen, Z. Li, Z. Song, Y. Liu, W. Chen, A. Knoll, A survey of the four pillars for small object detection: multiscale representation, contextual information, super-resolution, and region proposal, *IEEE Trans. Syst. Man Cybern. Syst.* 52 (2022) 936–953, <https://doi.org/10.1109/TSMC.2020.3005231>.



**Jijing Chen** received his B.E. degree in Opto-electronic Information Science and Engineering from Nanjing University of Posts and Telecommunications, China in 2019. Now he is working to pursue his doctor's degree in Opto-Engineering in Tianjin University. His main research interest is photoacoustic imaging.



**Kaixuan Ding** is a graduate student in the School of Precision Instrument and Optoelectronic Engineering of Tianjin University. He focuses on signal processing and object detection technology.



**Yihan Pi** received his B.E. degree in electronic science and technology from Tianjin University, China in 2019. Now she is working in School of Precision Instrument and Opto-Electronics Engineering, Tianjin University. Her main research interests are ultrafast laser technology and photoacoustic imaging.



**Shoujun Zhang** is a PhD student at Tianjin University. His research interests are terahertz metasurface devices and phase change materials.



**Jiao Li** is an Associate Professor in Tianjin University. She received her Ph.D. degree in Biomedical Engineering from Tianjin University in 2013, and after that she joined the School of Precision Instrument and Optoelectronics Engineering of Tianjin University. From 2015–2016, she was a visiting scholar at the Helmholtz Zentrum in Munich, Germany. Currently, she is working on multi-scale optoacoustic imaging with translational applications to tumor biology, and advanced molecular imaging with various functional nanomaterials and nanotechnologies for tumor theranostics.



**Zhen Tian** is the Vice President of the School of Precision Instrument and Optoelectronics Engineering of Tianjin University, the ECE China Director, Georgia Tech Shenzhen Institute, Tianjin University, and a Professor in Tianjin University. He is the member of Youth Working Committee of China Instrumentation Society, the member of Tianjin Optical Society, and the member of Optical Society of America. He received his Ph. D. degree in optical engineering from Tianjin University in 2010. His research interests are mainly focus in the fields of terahertz photonics, terahertz meta-materials and devices, terahertz applications, and terahertz photoacoustic imaging.