5

AI Machine Vision System for Wafer Defect Detection

Dmitry Morits, Marcelo Rizzo Piton, and Timo Laakko

VTT Technical Research Centre of Finland Ltd, Finland

Abstract

Surface defects generated during semiconductor wafers processing are among the main challenges in micro- and nanofabrication. The wafers are typically scanned using optical microscopy and then the images are inspected by human experts. That tends to be a quite slow and tiring process. The development of a reliable machine vision-based system for correct identification and classification of wafer defect types for replacement of manual inspection is a challenging task, due to the variety of possible defects. In this work we developed a machine vision system for the inspection of semiconductor wafers and detection of surface defects. The system integrates an optical scanning microscopy system and an AI algorithm based on the Mask R-CNN architecture. The system was trained using a dataset of microscopic images of wafers with Micro Electro-Mechanical Systems (MEMS), silicon photonics and superconductor devices at different fabrication stages including surface defects. The achieved accuracy and detection speed makes the system promising for cleanroom applications.

Keywords: AI, machine vision, semiconductor wafer, defect detection, convolutional neural network, Mask R-CNN.

5.1 Introduction and Background

One of the main challenges in micro- and nanofabrication is the identification and classification of surface defects. The defects are unavoidably generated

during processes such as chemical-mechanical polishing, photolithography, etching, diffusion and ion implantation, oxidation, metallization, and others [1][2]. The increasing complexity and density of semiconductor devices leads to an increase of the number of surface defects and dictates stricter requirements for defect detection. For example, contamination particles harmless for some design rules at the same time could be critical as the device dimensions grow smaller. The defect criteria are also varying in different locations of devices: for example, defects in a movable part or in the hermetic bond-sealing frame of a MEMS device are usually more severe than in secondary structures. Figure 5.1 illustrates microscopic images with surface defects generated during the microfabrication of different superconductor and semiconductor devices. Typical types of defects include particles, photoresist spots, edge defects, scratches, etc. It becomes evident that defect detection is an extremely important procedure, especially at the critical areas of the devices.

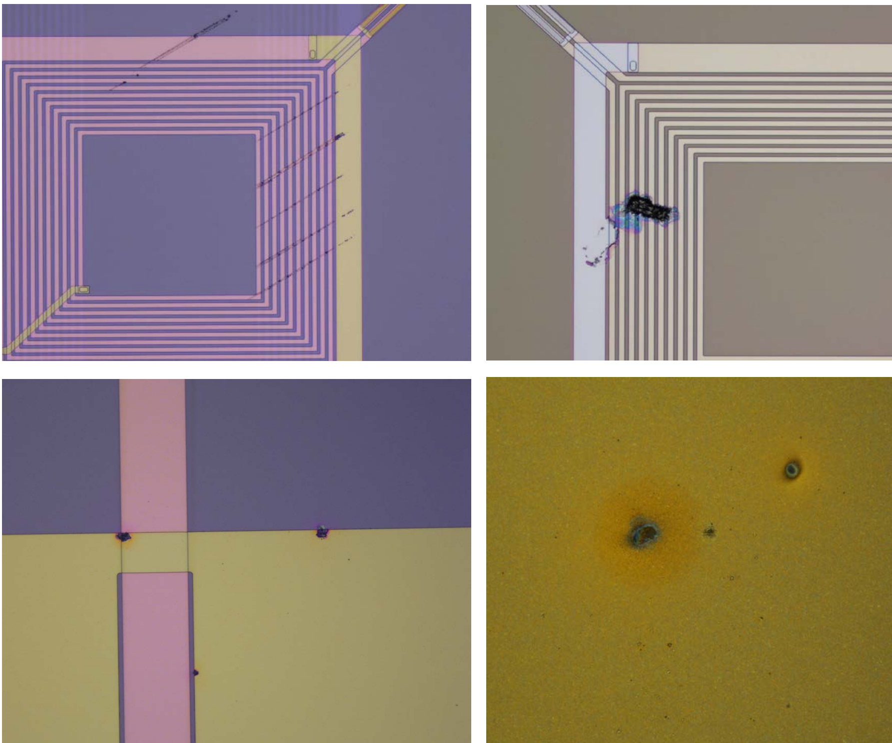

Figure 5.1 Examples of microscopic images of various superconductor and semiconductor devices with surface defects

VTT Micronova semiconductor fab is a Finnish national research infrastructure for micro-, nano- and quantum technology. The research areas include MEMS, photonic, quantum and other specialty components that can be used to create a wide range of sensors and devices. At VTT, the current visual inspection process of the wafer surface is manually performed by human experts. The wafers are scanned using optical microscopy, and then the images are inspected by the human experts. Since the inspection task requires extreme concentration, the time that an expert can perform the task is quite limited. Additionally, it tends to be a quite slow, tiring process and susceptible to human mistakes. Identification of defects by experts alone can potentially result in false identifications due to fatigue and lack of objectivity. The goal of this work is the development of a reliable machine vision-based system for the correct identification of wafer defects in the hope of replacing manual inspection. Moreover, this system would be directly integrated in the wafer inspection production line. Such a system would speed up the defect inspection, simplify the analysis and eventually help to improve the fabrication yield.

5.2 Machine Vision-based System Description

The general architecture of the developed machine vision system is shown in Figure 5.2. The wafers are inspected by a semi-automatic microscopy scanning system. In this work we tested both IJ 13 IR-inspector and Muetec CD3000 optical scanning system. The system produces a set of microscopic images, covering the full area of the wafer.

For the training of neural networks, we prepared an image dataset using microscopic images of wafers with MEMS, silicon photonics and superconductor devices at different fabrication stages including surface defects. The initial set included images of different resolutions and magnifications. First, we manually labelled the defects on each image and then cropped the areas with defects. The cropping allowed the increase of the dataset size and provided faster and more consistent training. Next, a data augmentation technique was used to increase the amount of data by adding slightly modified copies of already existing data, or newly created synthetic data from existing data. That procedure acts as a regularizer and helps to reduce overfitting when training a machine learning model [3]. In this case, the augmentation included mirror and rotation image transformation, as well as a change of the RGB spectre of the images. The full procedure of dataset preparation is

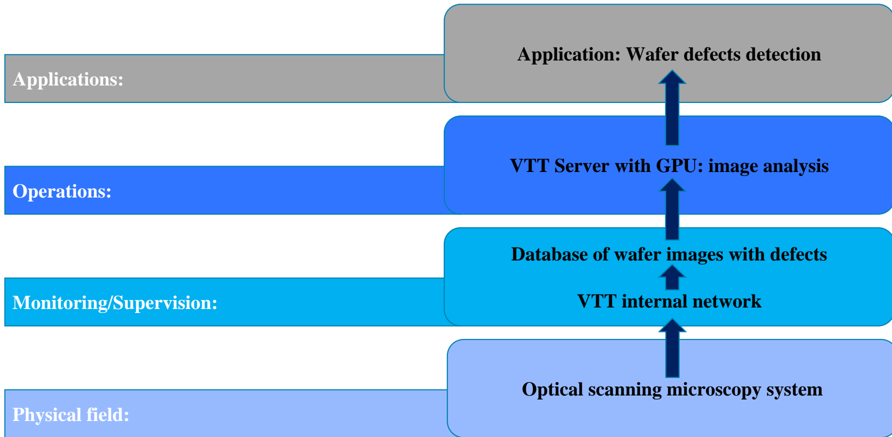

Figure 5.2 General architecture of the developed machine vision system

schematically shown in Figure 5.3. The dataset was split into training and validation sets, containing 935 and 165 images each.

Here we used a Convolutional Neural Network (CNN): a special type of deep learning algorithm, used primarily for image recognition and processing. CNNs are inspired by the organization of the animal visual cortex [4][5] and are designed to learn spatial hierarchies of features, from low- to high-level patterns. We developed an algorithm based on the Mask R-CNN architecture [6], which is a state-of-the-art algorithm for object detection - a computer vision technique that enables the identification and location of objects in an image or video. Mask R-CNN is the latest stage of evolution of CNNs, providing high detection accuracy. At the same time, it requires more computational resources compared to faster algorithms, such as YOLO [7]. Mask R-CNN consists of two stages. The first stage, called a Region

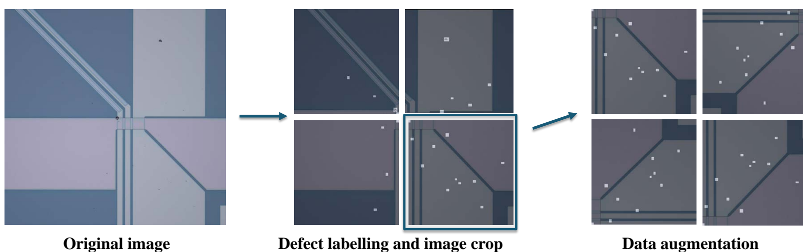

Figure 5.3 A scheme of the image dataset preparation, including labelling, cropping and data augmentation

Proposal Network, proposes candidate object bounding boxes. The second stage extracts features using Region of Interest Pool from each candidate box, then performs classification and bounding-box regression and outputs a binary mask for each Region. The ResNet-101 [8] convolutional backbone architecture was used for feature extraction over an entire image. The algorithm was optimized for so-called binary classification, which provides results in "defect vs background" format, without classification of defects, shown in Figure 5.4. The general comparison of the algorithm's performance to other object detection algorithms can be found in Refs [6] and [9].

Among the main requirements for the system are the functional suitability for defect detection, the integration of the scanning optical microscope and the server with the AI software, the usability for cleanroom users who are not familiar with the details of implementation, and the readability and visualization of the detection results for the users. The main KPIs for the system were: detection accuracy, time of processing a single image and evaluation by the cleanroom users from the points of usability and result readability. The AI algorithm based on the Mask R-CNN architecture passed several rounds of optimization and testing using microscopic images of various microelectronic devices.

There has been a significant progress in the application of deep learning techniques for wafer defect detection and classification [10]. The main innovation elements of this work compared to the state of the art is the integration of the algorithm with the scanning microscopy system, and training of the system using the dataset containing images of various devices at different stages of processing, instead of standard image databases available online. It allows the system to better distinguish between wafer defects and features of the devices and provides reliable detection of wafer defects for a wide range of semiconductor components.

To improve the system usability for the end-users, we implemented a Graphical User Interface adapted for cleanroom personnel not familiar with AI systems. The software was installed on a PC/server with NVIDIA Quadro RTX 5000 16GB GPU at the VTT Micronova cleanroom. Then the algorithm was integrated with the optical scanning microscopy system Muetec CD3000 by connection through the internal network. To improve the readability of the results, the system provides binary classification defect vs background with results available in both graphical and text formats. The feedback from the cleanroom experts helped in the improvement of system usability after several iterations of optimization. The testing results at the latest dataset with 192 images of 1600x1200px resolution and 5x optical magnification,

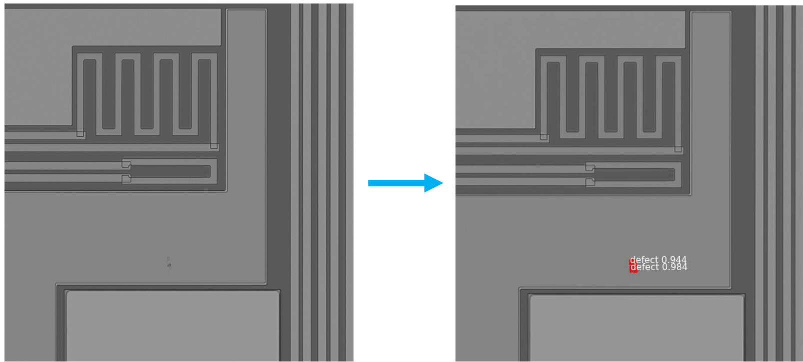

Figure 5.4 Example of binary classification of wafer defects: defect vs background

demonstrated 86% accuracy with a detection time of 1÷2 seconds per image. The accuracy of the system is approximately on the same level as that of a human operator, although it also depends a lot on the experience of the operators and their tiredness. The experts estimated 86% accuracy as sufficient for applications at VTT cleanroom but mentioned that only about 15% of the detected defects were critical for wafer processing. Unfortunately, the criteria of a defect being critical or non-critical is very device-specific and cannot be easily generalized. After the system provides the detection results, the final decision on the importance of the defects for processing had to be made by the cleanroom experts.

Regarding the system scalability, in the current work we did not have the goal of moving towards smaller technology nodes, although such scaling might require utilization of faster neural networks, like one-stage YOLO detectors. In general, the main expected impact of the system development is the reduction of the overall working time required for wafer defect inspection. We believe that the system will help saving valuable working time of cleanroom experts, improve fabrication yield and reduce fabrication cost.

5.3 Conclusion

We developed a system for the detection of wafer surface defects. The system integrates an optical scanning microscopy system and an AI algorithm based on the Mask R-CNN architecture. The image dataset used for training and testing the system included microscopic images of wafers with MEMS, silicon photonics and superconductor devices at different fabrication stages

including surface defects. The system demonstrated functional suitability for defect detection, high accuracy, and reasonable detection speed, making it suitable for potential cleanroom applications.

Acknownledgements

This work is conducted under the framework of the ECSEL AI4DI “Artificial Intelligence for Digitising Industry” project. The project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 826060. The JU receives support from the European Union’s Horizon 2020 research and innovation programme and Germany, Austria, Czech Republic, Italy, Latvia, Belgium, Lithuania, France, Greece, Finland, Norway.

References

[1] H. J. Queisser, E. E. Haller, “Defects in Semiconductors: Some Fatal, Some Vital”, Science, 281, 945– 950, 1998.

[2] T. Yuan, W. Kuo, and S. J. Bae, “Detection of spatial defect patterns generated in semiconductor fabrication processes”, *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 392–403, Aug. 2011.

[3] A. Buslaev, V. I. Iglovikov,; E. Khvedchenya, A. Parinov, M. Druzhinin, A. A. Kalinin, “Albumentations: Fast and Flexible Image Augmentations”, Information 11, 2, 2020. https://www.mdpi.com/2078-2489/11/2/125.

[4] S. Albawi, T. A. Mohammed, S. Al-Zawi, “Understanding of a convolutional neural network”, *International Conference on Engineering and Technology (ICET), IEEE*, pp. 1-6, 2017.

[5] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F.E. Alsaadi, “A survey of deep neural network architectures and their applications”, *Neurocomputing*, 234, pp. 11-26, 2017.

[6] K. He, G. Gkioxari, P. Dollár, R. Girshick, “Mask R-CNN”. arXiv:1703.06870, 2018.

[7] M. Carranza-García, J. Torres-Mateo, P. Lara-Benítez, J. García-Gutiérrez, “On the Performance of One-Stage and Two-Stage Object Detectors in Autonomous Vehicles Using Camera Data”, Remote Sens. 13, 89, 2021.

[8] K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learning for Image Recognition”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778, 2016.

[9] Z. Zhao, P. Zheng, S. Xu, X. Wu, “Object Detection With Deep Learning: A Review”, *IEEE Transactions on Neural Networks and Learning Systems*, 30, 11, 2019.

[10] U. Batool, M. I. Shapiai, M. Tahir, Z. H. Ismail, N. J. Zakaria, A. Elfakharany, “A Systematic Review of Deep Learning for Silicon Wafer Defect Recognition”, *IEEE Access*, 9, 116573, 2021.