Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 6
ISET 2024

Semiconductor Wafer Defect Detection Based on Machine Learning

Wenlong Gan*

School of Space Science and Physics, Shandong University, Weihai, China

* 201800830020@mail.sdu.edu.cn

Abstract. Semiconductor wafers are widely used in integrated circuit chips. Because of the complex production process, it is easy to cause various defects. Therefore, the defect detection of semiconductor wafers is an important means to ensure their yield and productivity. This paper mainly expounds on the detection method of wafer defects combined with the machine vision algorithm, including the CNN model, and the classification of the learning-based methods. Current problems and future prospects for total crystalline circle defect detection are discussed in the end.

Keywords: Semiconductor Wafer; Defect Inspection; Machine Learning.

1. Introduction

With the progress of semiconductor processing technology, semiconductor wafers have become more and more dense and complex, and the number of defects has greatly increased. In order to determine their root cause and eventually contribute to higher-quality products, these flaws should be found early on and accurately defined. This paper describes the classification of semiconductor wafer defects and the causes of their formation. And use machine learning and deep learning to detect and identify defects.

Several common wafer defects are described below, along with their causes. A poor workplace environment may cause particles and debris to enter the wafer surface, creating random defects. Enhancing the accuracy and stability of the production process can help lower the quantity of products that are deemed unqualified due to these kinds of flaws. System defects may be caused by factors such as the manufacturing process, equipment problems, or material quality. Repeatability, periodicity, and regularity in the wafer maps. In semiconductor manufacturing, combined defects are the most prevalent type of defects. They are often a combination of random defects and system defects[1]. According to the defect characterization and characteristics as the classification standard, three categories of wafer defects may be distinguished: morphological defects, contaminants, and crystal defects. The morphology defects of the wafer include poor flatness, edge cracks and surface defects, which can affect the formation of the IC pattern. This defect is usually caused during polishing or lithography by workers or machinery. Common pollutants, such as particles, are usually caused by dirty contamination of the wafer in the environment. At the molecular level, they include pollutants such as organic layers and inorganic layers. At the atomic level, including ions, heavy metal defects and so on[2].

Crystal defects usually result from heterogeneous crystal growth. Crystal defects at the level of the silicon wafers can affect the electrical performance of the devices built on these wafers. Even extreme degradation situations caused by crystal defects can lead to equipment failure.

At present, the detection methods for semiconductor wafer surface defects mainly include: artificial visual detection, optical detection, ultrasonic detection, and machine vision detection. The artificial visual method is performed by the operator using a microscope or naked eye to check the wafer surface for obvious defects. However, the efficiency of this detection method and the detection accuracy for small defects are low. The optical detection method uses a light source to illuminate the wafer surface. The defect is determined by detecting the intensity and shape of the reflected light and the transmitted light. This method belongs to non-contact detection and is suitable for defect detection in a large area. However, the detection effect of small and shallow defects is poor, which cannot detect the internal defects of the wafer. The ultrasonic detection method is to use the propagation speed, defect reflection and scattering characteristics of the ultrasonic wave in the wafer to detect its internal defects, such as cracks and cavities. However, this method is easily affected by the material density and sound speed, and the detection effect is worse in the specific material. In addition to the above three methods, another method is the machine vision detection method. This method utilizes image processing and computer vision techniques to detect surface defects on the wafer. It has the advantages of a non-contact detection process, high speed, automation, low cost, and scalability. Therefore, the detection method utilizing machine vision has turned into the main application method within the domain of wafer defect detection[3].This method is first optically examined by optical reflection, transmission, and interference. Images of the wafer surface were then acquired using a high-resolution optical microscope. Then the collected image undergoes graying, filtering, edge detection, image enhancement, feature extraction and so on. Then, the defects on the wafer surface are automatically identified and classified by the detection system. For each defect, assessment parameters such as its size, location, and severity are provided. Finally, according to the evaluation parameters of the defects, a defect report is generated to determine the cause of the defect.

2. CNN model

In computer vision, CNN is frequently used to extract translation-invariant features. CNN comprises the hidden layer, input layer and output layer. The essential component of CNN is the hidden layer, which typically comprises a convolutional layer, a pooling layer, and a fully connected layer. The representative high and low frequency characteristics in the image are extracted by CNN's front portion. More general and comprehensive characteristics are taken from the concealed, deeper layer. The following three characteristics demonstrate the unique qualities of the CNN-based approach. (1) The filter's sensitive reaction to the local input is guaranteed by the local connection. (2) The number of parameters is significantly decreased by weight sharing. (3) Pooling keeps important information while reducing the dimensionality of the data[4].

One of the CNN models for picture categorization applications is VGG-16[5]. It offers the benefits of great accuracy and versatility. It can address the issue of picture classification and location in hundreds of categories and process high-dimensional data fully. However, huge data samples are needed to support the VGG-16 training process. It also calls for a large amount of memory and processing power.

A modified and lightweight CNN model is called Mobilenet-v2. The lightweight model aims to address 2 issues for CNN[6]. (1) Many weight factors are included in hundreds of layer networks, necessitating a large amount of memory on the device to store them. (2) The speed standard in real life is frequently measured in milliseconds. Therefore, the computation must be decreased in addition to increasing processor performance. Mobilenet-v2 is ideal for implementation on mobile devices that emphasize memory utilization since it may lower parameters without sacrificing network performance.

Table 1. Detection Accuracy of WDD-Net and Other Methods

Model	Redundant	Crystal defects	Mechanical damage	Defect-free	Whole test set
VGG-16	99.86%	100%	100%	100%	99.96%
MobileNet-V2	100%	100%	100%	100%	100%
WDD-Net_28*28	100%	99.63%	97.52%	100%	99.70%
WDD-Net_224*224	100%	99.63%	94.63%	100%	99.44%

Table 2. Parameters and Model Size of WDD-Net and Other Methods

Model	Total parameters	Trainable parameters	Non-trainable parameters	Hdf5 Model size
VGG-16	133561036	133561036	0	1.50GB
MobileNet-V2	6030660	6008900	21760	46.2MB
WDD-Net	17200	17200	0	307KB

The detection speed of a new model, WDD-Net, aims to increase the model size and is further reduced by XIAOYAN CHEN et al. Wafer defect characteristics are not very complicated, hence the extraction of shallow characteristics is given more focus. WDD-Net has a simpler network structure than MobileNet-v2 and VGG-16. The GAP layer took the place of the completely linked layer in WDD-Net. Additionally, the maximum pooling layer is not used. Its layers are rather shallow and it makes use of smaller convolution kernels[4].

3. Learning-based Method for Wafer Defect Detection

Learning-based wafer defect detection methods utilize conventional machine learning techniques to identify and categorize wafer defects. The learning methods adopted mainly include supervised learning, unsupervised learning, mixed learning, semi-supervised learning and transfer learning. This section mainly discusses supervised, unsupervised and semi-supervised learning[7].

3.1. Supervised Learning

Table 3. Comparison of Major Supervised Machine Learning Methods

Type	Method	Accuracy	Contributions/limitations
SVM	**PCA+SVM**	0.8953	The extraction method of the main warehouse has good performance, and its robustness needs to be improved
	An aggregated decision tree based on Radon[8]	0.9050	Based on the Random transformation, the integrated decision tree summarizes the branch prediction with good aggregation performance
	Generalized uncertain decision tree	0.9560	Limited to cluster, circular, repeat, and random type defect identification
LDA	**LNLDA[9]**	0.9050	The classification accuracy was high, verifying the validity and utility of the model
Ensemble learning	**Bagging, voting, boosting**	0.9980	It maintains high classification accuracy for ring, spot and moon types
	AdaBoost Prediction	0.9400	Cut the expense of the test, improve the accuracy and efficiency
	SMBO	0.8130	Single-feature selection has better stability
	AdaBoost, bagging, RF, GBM	0.9645	Significantly reduce the test cost and unreliability of wafers
	LR, GB, RF	0.9587	Reduce the overfitting and under-fitting, and improve the classification performance

Supervised learning learns category features by using labeled training data to enable classification or prediction tasks. Providing strong support for decision-making, supervised learning can excel at automatically solving many complex problems.

3.2. Supervised Learning

Table 4. Comparison of Major Unsupervised Machine Learning Methods

Method	Accuracy	Contributions/limitations
DDPfinder[10]	0.9980	Accuracy increases with the running time and dataset, but also with the computational complexity
LCGMM+PC	0.9975	Effectively identify the wafer surface defect cluster, but the identified shape is simple
SDAE	0.9850	It is robust to noise with high precision
CAE	0.9929	Significantly improves the recognition accuracy
DBN	0.9120	Simplify WBM pattern extraction and improve the efficiency and accuracy of pattern diagnosis
MSA+PN	0.9232	Classification accuracy is highest among small sample methods, but can only identify a limited number of defect categories

The purpose of unsupervised learning is to discover the underlying structure and characteristics of the data from label-free data without giving a target label. Compared with supervised learning, unsupervised learning does not rely on manually labeled data for training but relies on the algorithm to learn and explore the data independently.

3.3. Semi-supervised Learning

Table 5. Comparison of Major Semi-supervised Learning Methods

Method	Accuracy	Contributions/limitations
SCSDAE	0.9883	Good robustness and feature extraction performance have overfitting problems
ESDAE	0.9703	Using the manifold regularization technique, the algorithm works well
PCACAE	0.9377	Training cost is low but poorly applicable, can only identify single defect patterns and cannot handle nonlinear data
Bagging+SOM+CNN	0.9437	Classification accuracy and high efficiency
DBSCAN+SOM	0.9437	Solve the problem of marker scarcity and error, improve the effectiveness of the prediction model, and obtain a higher pattern recognition rate
LADDER networks	0.9020	High precision classification of GFA mode by identifying the root cause of the total failure area GFA
SVAE[11]	0.9030	WBM with the total failure area GFA mode with high accuracy
SS-CDGMM	0.9488	Classification accuracy and high efficiency
SS-AIR	0.9395	It is more robust to random defects and requires due to noise filtering algorithm, and is unable to learn new defect patterns independently

Due to the potential supervised learning constraints, limited by the quantity of labels that are available and the classification performance of unsupervised learning on unlabeled defects, semi-supervised learning has emerged.

Combining semi-supervised learning with deep generative models provides a novel approach to dealing with mixed-type defect patterns. By introducing multiple latent class variables, the model is able to more precisely capture the differences between different patterns, providing a more accurate representation of the classification problem. The application of this approach promises remarkable achievements in the field of wafer defect detection as well as in other complex pattern classification problems[14].

4. Prospection

There are still many aspects of machine learning technology that need improvement, such as the high requirements for data sets and the high cost of collection of data collection.

Due to the time cost and difficulty of collecting the wafer defect dataset, the algorithm lacks sufficient samples to learn the characteristics and change patterns of the defect during training.

The wafer defects may have different scales and dimensions. For defects on the same scale, there are changes in shape and texture. These uncertainty factors can influence the characteristic representation of the defect.

5. Conclusion

CNN models such as VGG-16 and MobileNet-V2 perform excellently in wafer defect recognition accuracy, but they have a complex structure with numerous parameters and models. WDD-Net simplifies the structure and minimizes the quantity of parameters and computations but decreases the identification accuracy marginally. Although the WDD-Net model is slightly inferior to MobileNet-V2 and VGG-16 in terms of identification accuracy, WDD-Net is compact with few parameters and a small model size.

In the wafer defect detection, according to the different learning methods, it is primarily separated into several categories. Supervised learning is able to perform classification or prediction tasks and performs well at automatically solving complex problems. Compared with supervised learning, unsupervised learning does not rely on manually labeled data for training, and the data can be learned and explored autonomously through algorithms. Semi-supervised learning uses labeled wafer maps to learn relevant features of each defect pattern. Meanwhile, it utilizes unlabeled wafer maps to further refine the feature representation.

References

[1] Kaempf U. The binomial test: A simple tool to identify process problems[J]. IEEE Transactions on semiconductor manufacturing, 1995, 8(2): 160-166.

[2] Bollmann W. Crystal defects and crystalline interfaces[M]. Springer Science & Business Media, 2012.

[3] Tang B, Kong J, Wu S. A review of machine vision surface defect detection[J]. Chin. J. Image Graph, 2017, 22(12): 1640-1663.

[4] Chen X, Chen J, Han X, et al. A light-weighted CNN model for wafer structural defect detection[J]. IEEE access, 2020, 8: 24006-24018.

[5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[6] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.

[7] Lei L J, Sun S L, Xiang Y K, Sun Y and Liu H K. 2020. Bottleneck problem of computer vision application in intelligent manufacturing. Journal of Image and Graphics, 25(07): 1330-1343.

[8] Piao M, Jin C H, Lee J Y, et al. Decision tree ensemble-based wafer map failure pattern recognition based on radon transform-based features[J]. IEEE Transactions on Semiconductor Manufacturing, 2018, 31(2): 250-257.

[9] Yu J B, Lu X L and Zong W Z. 2016. Wafer surface defect detection and identification based on dynamic integration of local and nonlocal linear discriminant analysis and gaussian mixture model. IEEE/CAA Journal of Automatica Sinica, 42(1): 47−59.

[10] Taha K, Salah K, Yoo P D. Clustering the dominant defective patterns in semiconductor wafer maps[J]. IEEE Transactions on Semiconductor Manufacturing, 2017, 31(1): 156-165.

[11] Kong Y, Ni D. A semi-supervised and incremental modeling framework for wafer map classification[J]. IEEE Transactions on Semiconductor Manufacturing, 2020, 33(1): 62-71.