PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Detection and correlation of yield loss induced by color resist deposition deviation with a deep learning approach applied to optical acquisitions

Alcaire, Thomas, Le Cunff, Delphine, Tortai, Jean-Hervé, Soulan, Sebastien, Gredy, Victor, et al.

Thomas Alcaire, Delphine Le Cunff, Jean-Hervé Tortai, Sebastien Soulan, Victor Gredy, Maxime Templier, Mehdi Kessar, Raffaele Bianchini, Audrey Berthoud, "Detection and correlation of yield loss induced by color resist deposition deviation with a deep learning approach applied to optical acquisitions," Proc. SPIE 11611, Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV, 116112G (22 February 2021); doi: 10.1117/12.2582058

Event: SPIE Advanced Lithography, 2021, Online Only

Detection and correlation of yield loss induced by color resist deposition deviation with a deep learning approach applied to optical acquisitions

Thomas Alcaire*[a,b], Delphine Le Cunff[a], Jean-Hervé Tortai[b], Sebastien Soulan[b], Victor Gredy[a], Maxime Templier[a], Mehdi Kessar[a], Raffaele Bianchini[a], Audrey Berthoud[c]
[a]*STMicroelectronics*, 38920 Crolles ; [b]*Univ. Grenoble Alpes, CNRS, CEA/LETI-Minatec, Grenoble INP*, LTM*, F-38054 Grenoble-France; [c]*CEA-LETI Minatec*, F-38054 Grenoble

ABSTRACT

On imager devices, color resists are used as optical filters to produce RGB pixel arrays. These layers are deposited through spin coating process towards the end of the fabrication process flow, where complex topography can induce thickness inhomogeneity effect over the wafer surface causing a radial striations signature, predominant at the edge of the wafer. This deviation can induce important yield loss but is hardly detectable with standard inline metrology or defectivity approach. In this study, an interferometry-based metrology system and a reflectometry-based defectivity system were used to gather raw optical responses on the full wafer surface. Individual die cartographies were created from those and a deep learning algorithm was trained from both optical techniques. We then applied the deep learning algorithm on a specific set of test wafers to determine the number of dies affected by striations. From there, we evaluated the correlation of the outcome classification with the final electrical tests done on each die of those wafers.

Keywords: Predictive metrology, Model-less approach, Deep learning

1. INTRODUCTION

In recent years, the acceleration of data analysis developments enabled the introduction of new strategies in the metrology field aiming faster process control decision without the need of a model development. Model-based metrology techniques traditionally rely on a precise knowledge of the materials and architecture of target structures and necessitates an accurate calibration of the model that is very time and resource consuming. On the other hand, model-less approaches rely on the strength of artificial intelligence (AI) to compensate the absence of a rigorous model but generally requires a large training dataset. Here, we present a case of AI application for the detection of a process deviation namely thickness inhomogeneity at the wafer scale.

The deviation studied in this paper is related to the color resist depositions, introduced on imager devices acting as color filters above each pixel to create the RGB arrays. These resists are successively deposited by a spin coating process, but the complex surface topography present on the wafer at this step of the process fabrication can induce thickness inhomogeneity at large frequency scale. This leads to the apparition of a radial signature identified as “striations” over the full wafer, more prominent at the edge of the wafer. Considering the optical nature of the devices, this can strongly interfere with the optical performances and potentially result in a yield loss. This is generally secured and detected at the final optical testing step (called EWS for Electrical Wafer Sorting) but far too late for any process rework and correction on the wafers.

Unfortunately, this deviation is hard to properly detect with a traditional thickness metrology or defectivity approach. Indeed, inline metrology measurements are performed in dedicated blanket structures which size is small compared to the resist thickness frequency scale variation and they are generally located in scribe lines where surrounding topography is different than inside the device. Also, this deviations effect is stronger on the extreme edge of the wafer where only likely few metrology targets are accessible. Furthermore, most of the metrology in-line measurements rely on a rigorous model of the underlying layer stack. The color resists used have very complex optical properties, and modeling their optical responses, especially considering the underlying stack, would be extremely arduous.

Metrology, Inspection, and Process Control for Semiconductor Manufacturing XXXV,
edited by Ofer Adan, John C. Robinson, Proc. of SPIE Vol. 11611, 116112G
 · doi: 10.1117/12.2582058

Similarly, detecting striations would also be challenging for a traditional defectivity approach because this thickness variation is not by proper sense a physical defect (i.e. scratches, bridges…) that can easily be detected by a generic die-to-die or reference die compare approach. This approach would indeed be ineffective for this particular deviation because most dies are affected, to a different extent, by this deviation.

We envision that an approach using the strength of both approaches could then be more successful in detecting the striations. This is why we chose to use the sensitivity of the optical techniques used in metrology combined with the full wafer approach of defectivity, by gathering raw optical data on the full surface of wafers affected by the striation's deviation.

The goal of this work is to develop a model-less approach that uses two industrial optical-based systems to gather raw data on the full surface of wafers in order to detect the presence of striations inside the dies. We already evaluated such an approach in a previous work [1] where an academic spectroscopic ellipsometer was used to gather raw optical data on a full wafer scale. A machine learning algorithm successfully classified the resulting cartographies depending on whether they exhibited a deviations signature. With this work, we intended to use industrial equipments combined with a deep learning algorithm to propose a robust applicable approach.

The studied wafers are part of a R&D lot exploring different process flows and materials to deduce the optimal conditions refer to striations' effects and related yield loss. In order to obtain significant variations in the final yield results, intentionally sub-optimal conditions were chosen. The gathered optical data will be treated to generate single die cartographies, which will be used to automate the detection process with deep learning. The algorithm will be trained to classify the dies whether they exhibit a striations signature or not. Once trained, it will be used to find the ratio of striated dies of each wafer studied. Finally, we will correlate those results with the EWS results in order to develop a predictive model from those early optical measurements.

2. EXPERIMENTAL SETUP AND PROTOCOL

2.1 Wafer process description

In order to build imagers devices, color resists are used as optical filters to produce RGB color pixel arrays. They are laid on top of the image sensors (following a Bayer filter mosaic), allowing the reproduction of the image observed by the device by combining the signals obtained from each color. However, because each pixel is filtered to receive only one of the three colors, it is necessary to use a demosaicing algorithm, applied to surrounding pixels to estimate the actual RGB value. A representation of the pixel array and a summary of the demosaicing algorithm is show in Figure 1.

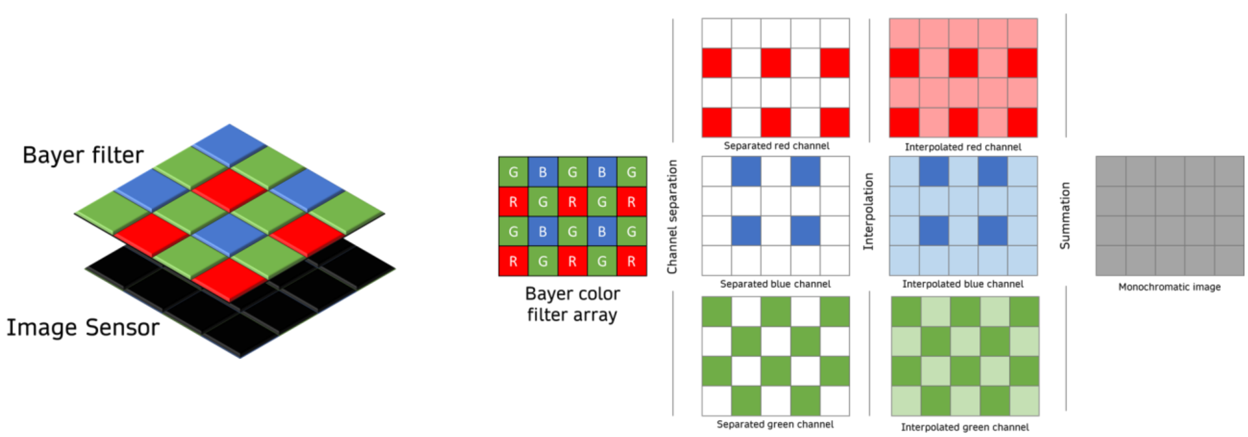

Figure 1 – Architecture of the imager devices pixel array and summary of the demosaicing algorithms.

The color resists depositions are performed close to the end of the process, where the wafers top surface can present nano-topography variation due to the presence of connection features patterned at previous process step. Since the resists are deposited through spin coating, this topography creates obstacles in the flow of the resist, sometimes leading to an inhomogeneous repartition of the resist thickness over the whole wafer surface. A striation signature can then be observed, characterized by radial waves, more pronounced the further from the center of the wafer. After the color resists are laid, micro-lenses are deposited on top of each pixel, in order to focus the light on the sensors. The thickness variation induced by the striations will then hinder the lens focalization, leading to optical yield loss and improper image reconstruction by the affected devices. Detecting this deviation and correlating its effect on optical yield loss is then of the utmost importance, because it will allow actionable decisions early in the process to correct the thickness inhomogeneity.

In order to identify the optimal process flow to reduce this deviation, a R&D 24-wafer lot was processed with varying conditions, resulting in largely different optical yield results in the final EWS tests. This will allow us to develop a predictive model with a sufficient range of yield. A simplified split matrix, showing each wafers process flow, is shown in Table 1. We used this lot to evaluate our model-less approach based on optical acquisitions to detect this deviation, and because these wafers would be fully processed and tested, we could correlate our findings with the final yields of each wafer. A 25th wafer was processed with the “usual” flow for a baseline comparison.

		Wafers																							
OPERATION	RECIPE	3	8	13	15	19	24	17	23	11	6	14	18	4	22	12	7	5	2	20	10	16	1	9	21
Process step 1	Old	x	x	x	x	x	x	x	x	x	x	x	x												
	New													x	x	x	x	x	x	x	x	x	x	x	x
Process step 2	Standard	x	x	x	x	x	x							x	x	x	x	x	x						
Process step 3	Standard	x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x	
Green deposition	Standard	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Blue deposition	Old	x	x	x				x	x	x				x	x	x				x	x	x			
	New				x	x	x				x	x	x				x	x	x				x	x	x
Red deposition	Standard	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Table 1 – Process split matrix intended to find the optimal process flow to reduce the striations deviation’s impact

2.2 Optical equipments description

Two optical-based systems were used during this study. The first one is a metrology equipment named PWG (Patterned Wafer Geometry). It relies on a double Fizeau interferometer to obtain information about the wafer’s topography. During the acquisition, the wafer is held vertically to remove the influence of gravity and both frontside and backside are analyzed. It is usually implemented in manufacturing lines for the control of stress, shape or nanotopography of product wafers. The acquisitions are performed using a 635nm laser, allowing a 100 µm spatial and 0.5 nm vertical resolution. Unfortunately, because the color resists studied are transparent at the given wavelength, the interferometry can sometimes lead to incorrect reconstruction of the wafer’s topography (less than 10% of dies impacted). Furthermore, this system is designed to measure local nanotopographies variations lower than 150nm. Consequently, this system cannot be used in an automatic industrial process control protocol to reliably detect striations signature on those imagers’ devices.

The second equipment is a defectivity system: ALTAIR, that uses reflectometry to detect defects on the surface of the wafer. It uses three LEDs as light sources (Blue at 450nm, Green at 520nm and Red at 625nm) that can be combined to obtain a white source. Two acquisition protocols are available: Brightfield and Darkfield, where the incidence is respectively normal and grazing, each specialized in detecting specific types of defects. This equipment is traditionally used with a die-to-die comparison approach in order to detect variation in signature from one die to another, then an automatic classification is performed to identify the detected defect. For such an approach, and using its multiples objectives, a 1.2 µm resolution can be reached, but for a full wafer acquisitions, the spatial sensitivity is lessened to 150 µm. A simplified schematic of each system is presented in Figure 2.

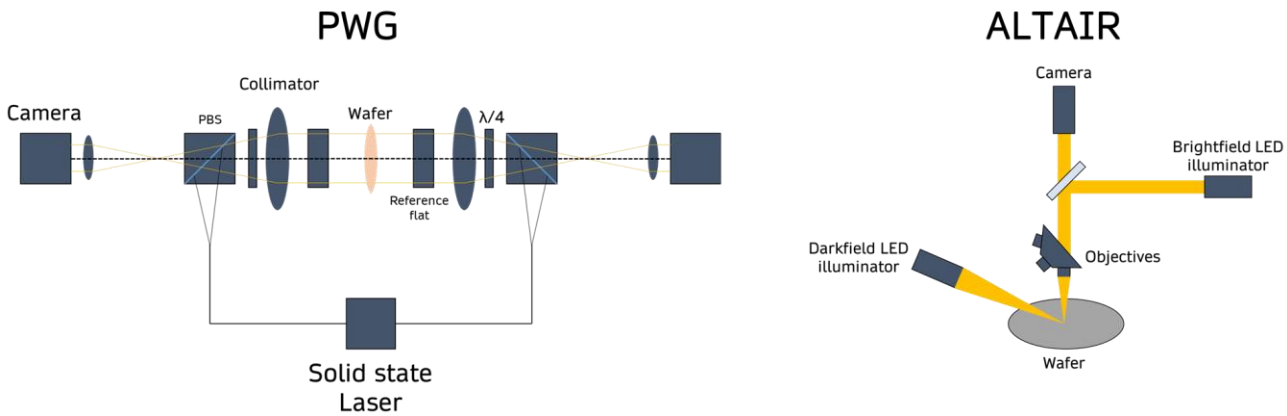

Figure 2 – Functional schematics for the PWG and ALTAIR systems that were used for the full wafer acquisitions.

Both acquisitions were performed after the last resist deposition when the RGB pixel array is completed. This would allow, when this approach is used in production, for actionable decisions on the concerned wafers, for example in the form of a rework of the depositions process step.

An example of a full wafer acquisition for both equipments is shown in Figure 3, where the striations signature is clearly visible. Our goal was to train a deep neural network to automatically identify each of these dies and classify them depending on the presence of a striations signature.

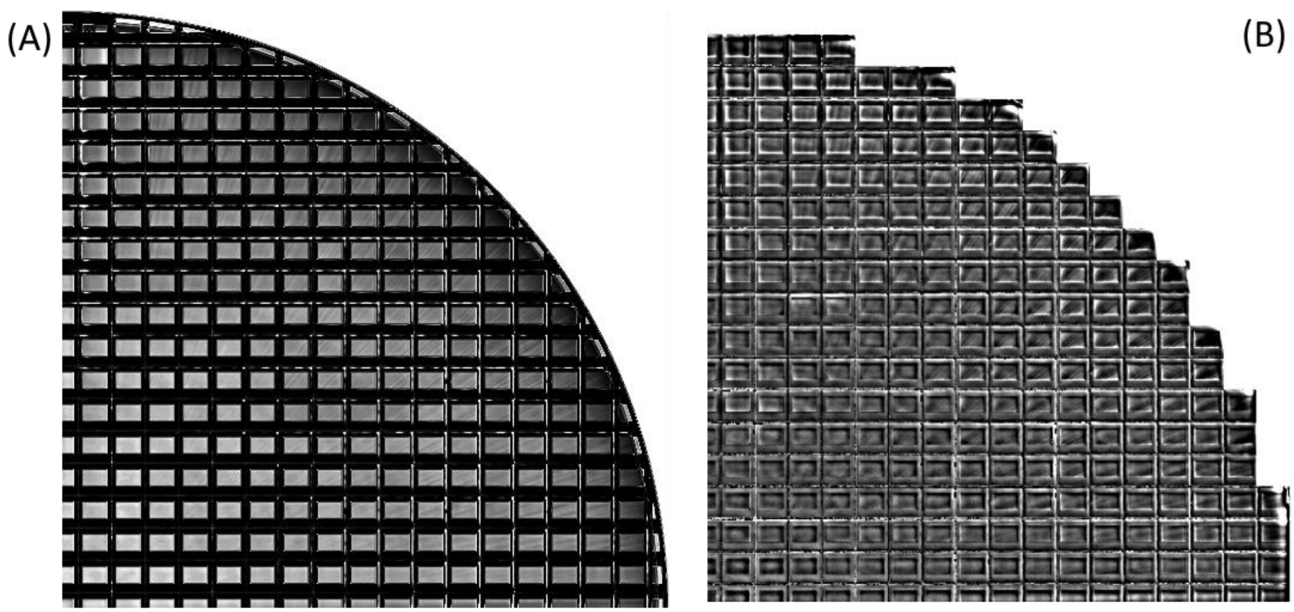

Figure 3 – Full wafer images obtained from the PWG (A) and the ALTAIR acquisitions (B) on a wafer affected by striations.

2.3 Deep learning methodology

For many years, deep learning algorithms have been used for image classification problems, but Convolutional Neural Networks (CNN) represents a huge breakthrough in this domain. They have been adopted as the main solution for image recognition and analysis in various domain such as medicine and self-driving cars. They are fast and efficient and only require a labeled dataset to be trained. From there, the deep learning network is able to automatically learn to identify each class. Due to their robustness, accuracy, and unmatched pattern recognition ability, such networks are used in a semiconductor industrial setting for fault detection and root cause determination [2], or wafer map defect pattern classification [3].

In this study, we used *Resnet-18* which is a CNN composed of 18 layers [4]. The structure of this network is shown in Figure 4. And, in order to reduce the amount of training data needed to train such a network from scratch, which will require thousands of identified dies, we used a transfer learning approach. This relies on using a neural network already trained to identify other objects (i.e. cars, animals…) and feeding it with another dataset to reset its classification capability, according to our problem. This way is considerably faster and less demanding in data quantity. This allowed us to use the R&D lot as a training data source while only using a very limited number of dies per wafer.

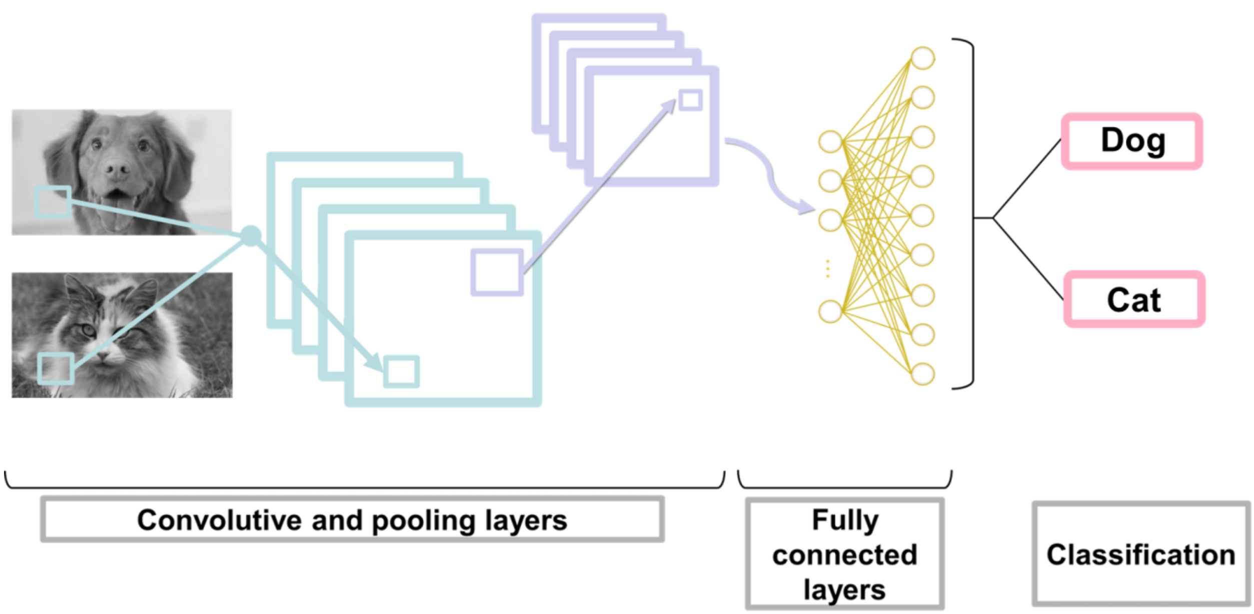

Figure 4 – Typical structure of a Convolutional Neural Network used to classify images into different classes. The convolutive and pooling layers transform the images to extract the discriminative features of each class, and the fully connected layers uses those features to perform the image classification.

3. RESULTS

3.1 Acquisition results and datasets creation

Once the RBG array process completed, we acquired he raw optical signals on the full surface of the 24 wafers of the R&D test lot as well as the base line wafer (25th) both on PWG and ALTAIR equipments. The data were used to generate individual cartographies of each dies. In order to automate the process, the raw data were treated by a custom-made Python script.

With ALTAIR acquisitions, we obtained, as output, full wafer images that needed to be divided into individual dies. This was done by using the wafer and mask information in order to obtain the coordinates and size of each die, and then creating sub-images from the full wafer source. With PWG, the output consisted of the raw nanotopography data for the 7 million of points measured on the full wafer surface. We used the previously mentioned wafer and mask information to generate images from the measurements for each die of the wafers studied.

In order to train and evaluate the neural networks, we provided a set of dies that we manually identified as s*triated*, and another one of dies that we identified as n*on-striated*. A third set of data was necessary for PWG data because of the aforementioned interferometry issues with transparent resists, that we identified as r*econstruction error*. The different datasets were made using the protocol presented in Figure 5, along with an example of each die cartographies class for both equipments.

	ALTAIR	**PWG**
CNN training dataset	10 dies × 2 class × 24 wafers → 480 dies	10 dies × 3 class × 24 wafers → 720 dies
CNN evaluation dataset	980 manually classified dies (not used in training – 25th wafer)	980 manually classified dies (not used in training – 25th wafer)

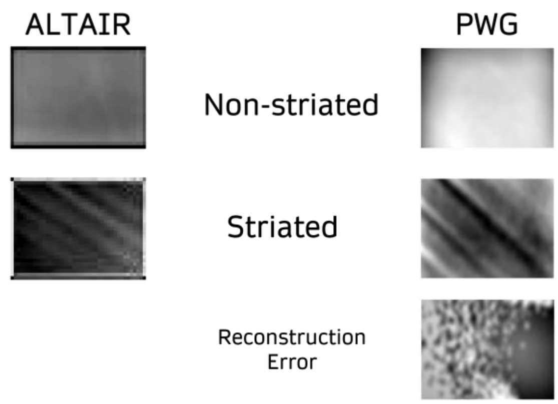

Figure 5 – Dataset creation protocol for the training and evaluation of the deep neural networks and example of each die class for both reflectometry technique (ALTAIR) and interferometry technique (PWG)

We used the previously mentioned dataset to train two convolutional neural networks, one for each equipment. The transfer learning approach allowed us to only use a fraction of our dataset to train the networks, considering each wafer is composed of around 1000 dies. Once trained, these networks were tested using the evaluation datasets. The CNN trained from ALTAIR data correctly identified each die at a rate of 96.42%, while the PWG neural network exhibits a slightly lesser accuracy of 93.05%, most likely because of the additional *reconstruction error* class. Both deep learning treatment were then highly effective in classifying our die cartographies based on the presence or absence of a striations signature. This allowed the automatic and accurate classifications of all the dies created from the 24-wafer lot PWG and ALTAIR acquisitions, that would have been extremely fastidious otherwise.

We compared the classification obtained by both algorithms to verify their concordance, as can be seen in Figure 6. As expected, most of the striated dies are located on the edge of the wafer. Furthermore, both networks predictions agree very well for most dies, the differences resulting mainly from reconstruction errors from PWG acquisitions.

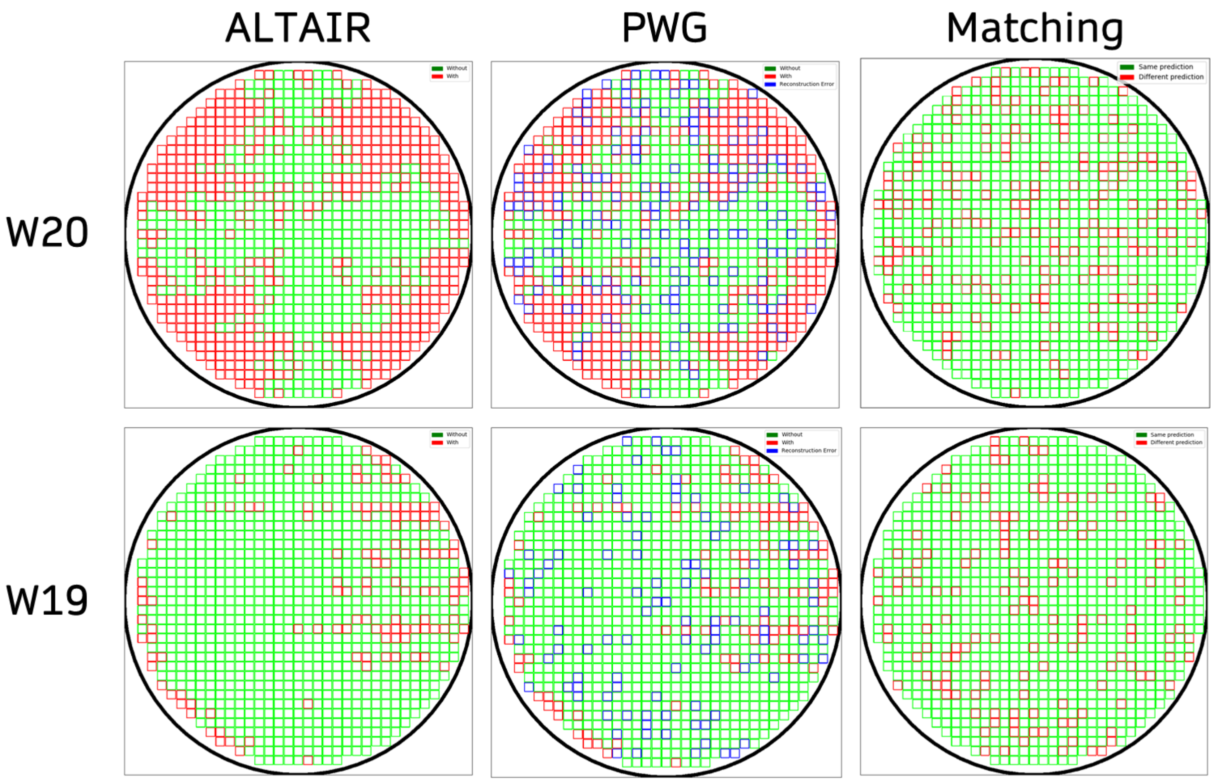

Figure 6 – Example of the repartition of dies classified by the neural network for each data source and matching between the predictions.

3.2 Deep learning treatment and EWS Correlation

Once the die cartographies were automatically classified by the neural networks, we were able to calculate the proportion of dies affected by the striation's deviation, for each wafer studied and to compare with the final optical yield results of the studied lot. Table 2 reports the final yield results for each wafer are presented in along the predicted percentage of dies affected by striations by the neural networks for both instruments. Before other considerations, we can see that both systems are in agreement with the final optical test when it comes to defining the best process route. The 15/19/24 sub-lot is the best performing in terms of optical yield, and also the one with the smallest percentage of striated dies determined by the CNNs trained on ALTAIR and PWG data.

	Wafers																							
OPERATION	3	8	13	15	19	24	17	23	11	6	14	18	4	22	12	7	5	2	20	10	16	1	9	21
Final optical test (Normalized % of defective dies)	35.2	73.0	88.1	6.8	0.0	2.9	64.4	79.6	69.0	6.8	20.1	35.2	96.9	94.9	86.9	15.7	14.4	1.6	97.6	98.9	100.0	15.7	14.4	13.1
ALTAIR – Deep Learning predictions (Normalized % of striated dies)	29.3	72.9	78.1	0.0	3.4	0.5	59.6	72.0	58.9	12.9	9.9	18.1	79.9	86.7	88.5	7.0	15.1	6.1	86.7	94.6	100.0	25.5	24.8	24.4
PWG – Deep Learning predictions (Normalized % of striated dies)	55.8	86.0	100.0	10.0	9.2	0.0	81.7	92.2	86.5	26.7	33.4	29.9	93.3	90.3	87.1	19.4	17.5	14.0	88.1	91.9	92.5	9.7	8.1	9.2

Table 2 – Final optical results for each wafer of the R&D lot that will be used to correlate with our neural network predictions, with corresponding percentage of dies identified as *striated* by the deep learning algorithm, for each measurement system.

The correlation results are presented in Figure 7. We observe a strong correlation for both techniques, considering a large variation in the yield result. The PWG correlation is slightly worse, most likely because we ignore the dies identified as *Reconstruction Error*, inducing a smaller statistical amount of dies considered.

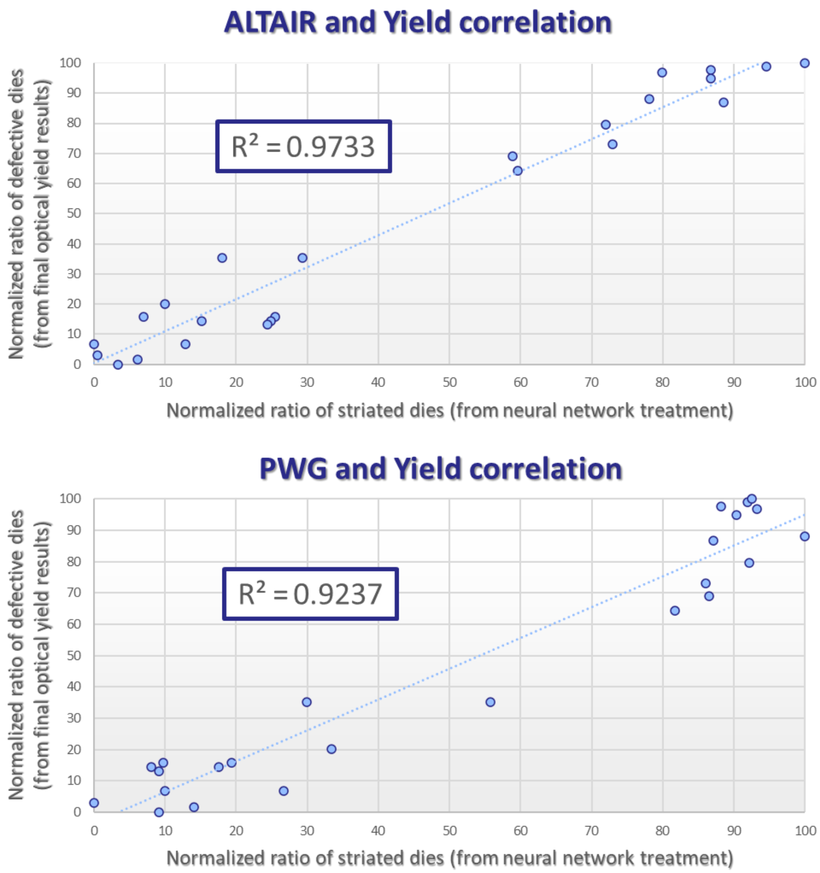

Figure 7 – Correlation between the proportion of striated dies, determined by the CNN treatment, and the proportion of defective dies, established with the final optical tests, for the ALTAIR and the PWG acquisitions on the 24-wafer lot.

A hybrid approach, whose goal was to take advantage of both equipments capabilities, was tested on this deviation. Both ALTAIR and PWG cartographies dies were stacked, and a CNN was trained to classify the resulting hybrid cartographies. Unfortunately, it was unsuccessful because of interferences between the systems responses, where a striations peak would result in a high response on a nanotopography measurement, but on a low one in a reflectometry measurement. This induced an overall fading of the striations signature, leading to poor correlation results.

4. CONCLUSION

Using metrology and defectivity systems to gather optical responses early in the process, we successfully detected a full wafer scale deviation. The gathered data was used to generate die cartographies that would be automatically classified by a deep learning algorithm. This classifications results proved a correlation between the proportion of dies affected by the striation deviation and the final electric results of the wafer. In order to validate our approach, a predictive model created from the correlation done in this work will be used and compared to the actual yield obtained in the final optical tests of production lots. This will allow the early process deviation detection and correction on the production line.

ACKNOWLEDGMENTS

This work has received funding from the Electronic Component Systems for European Leadership Joint Undertaking under grant agreement No 826589 (project MADEin4). This Joint Undertaking receives support from the European Union Horizon 2020 research and innovation program and, Austria, Belgium, France, Germany, Hungary, Ireland, Italy, Netherlands, Romania, Sweden and Israel.

REFERENCES

[1] Alcaire, Le Cunff, Gredy, Tortai, "Spectroscopic Ellipsometry Imaging for Process Deviation Detection via Machine Learning Approach," Proc. ASMC 2020, 1-6 (2020).
[2] Lee, Ki, Cheon, Sejune, Kim, Chang Ouk, "A Convolutional Neural Network for Fault Classification and Diagnosis in Semiconductor Manufacturing Processes," Proc. IEEE 2017, 1-1 (2017).
[3] Nakazawa, Takeshi, Kulkarni, Deepak, "Wafer Map Defect Pattern Classification and Image Retrieval using Convolutional Neural Network," Proc. IEEE 2018, 1-1 (2018).
[4] He, Kaiming, X. Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition," Proc. IEEE CVPR 2016, 770-778 (2016).