Virtual overlay metrology for fault detection supported with integrated metrology and machine learning

Hong-Goo Lee [a], Emil Schmitt-Weaver[*] [b], Min-Suk Kim [a], Sang-Jun Han [a], Myoung-Soo Kim [a], Won-Taik Kwon [a], Sung-Ki Park [a], Kevin Ryan [b], Thomas Theeuwes [b], Kyu-Tae Sun [b], Young-Wan Lim [b], Daan Slotboom [b], Michael Kubis [b] & Jens Staecker [b]

a) SK hynix, Korea
b) ASML Netherlands B.V., De Run 6501, 5504 DR Veldhoven, the Netherlands

ABSTRACT

While semiconductor manufacturing moves toward the 7nm node for logic and 15nm node for memory, an increased emphasis has been placed on reducing the influence known contributors have toward the on product overlay budget. With a machine learning technique known as function approximation, we use a neural network to gain insight to how known contributors, such as those collected with scanner metrology, influence the on product overlay budget. The result is a sufficiently trained function that can approximate overlay for all wafers exposed with the lithography system. As a real world application, inline metrology can be used to measure overlay for a few wafers while using the trained function to approximate overlay vector maps for the entire lot of wafers. With the approximated overlay vector maps for all wafers coming off the track, a process engineer can redirect wafers or lots with overlay signatures outside the standard population to offline metrology for excursion validation. With this added flexibility, engineers will be given more opportunities to catch wafers that need to be reworked, resulting in improved yield. The quality of the derived corrections from measured overlay metrology feedback can be improved using the approximated overlay to trigger, which wafers should or shouldn't be, measured inline. As a development or integration engineer the approximated overlay can be used to gain insight into lots and wafers used for design of experiments (DOE) troubleshooting. In this paper we will present the results of a case study that follows the machine learning function approximation approach to data analysis, with production overlay measured on an inline metrology system at SK hynix.

Keywords: machine learning, overlay, fault detection, virtual metrology, integrated metrology

1. INTRODUCTION

With the case study "Overlay improvements using a real time machine learning algorithm" [1] the authors of this paper describe how a neural network used for function approximation can be trained to pickup on unique and dynamic process instabilities as they contribute to the on product overlay budget. The neural network in this study was trained by pairing input (wafer metrology and context from a TWINSCAN system) to output (overlay metrology coming from an integrated YieldStar system). For the theoretical study [1] our goal was to show that exposure side overlay could be improved by training a function to pickup on pre exposure cues within the data. With positive proof book results the logical next step would be to realize this on real wafers within the fabrication environment. From the point of view of the scanner software, this means we need to be able to make small changes to the exposure side correction strategy after each wafer is clamped and receives pre exposure metrology (alignment and leveling).

Understanding the time and resources required to make low level changes in the TWINSCAN system software, the authors of this paper have reimagined the application of the neural network used for function approximation in the case study [1]. The idea being that instead of predicting small changes in overlay in real time we take a retrospective approach and look at predicted overlay vector maps for all of the wafers coming off the track. Then by applying the same POR (process of record) methodology to the predicted overlay as is used for monitoring overlay metrology, we can easily see if any of the wafers in the lot standout of the standard population. For a process engineer, wafers or lots with predicted overlay signatures outside the standard population could see a few potential actions. If the wafers were not measured

[*] Correspondence: emil.schmitt.weaver@asml.com

Metrology, Inspection, and Process Control for Microlithography XXIX, edited by Jason P. Cain, Martha I. Sanchez, Proc. of SPIE Vol. 9424, 94241T · © 2015 SPIE
CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2085475

with an integrated YieldStar system, they can be redirected to offline metrology for excursion validation, potentially improving yield. Or if the wafers were measured with the integrated YieldStar system, the behavior of the predicted overlay signature of the wafer measured with respect to other wafers in the same lot can be use to judge if the wafer should be use for APC (automated process control). For a development or integration engineer, the predicted overlay can bring invaluable insight into setting up or fine tuning a process. As we will show in the Results (Section 3) of this paper, when multiple lots are used for a DOE (design of experiment) the user can more definitively converge on an optimum process setting when all the wafers in the lot have a predicted overlay vector map.

2. BUILDING THE PREDICTIVE APPLICATION

2.1 Assembling database for training and testing

To train and test our function we only look at the wafers from the lots that have YieldStar metrology. These are the wafers we have both input and output data for. The wafers are randomly selected to be in the training and testing groups (see 2a and 2b in Figure 1). The wafers from testing group 2b in Figure 1, green wafers, are set aside and will not be used in any capacity for the training step. The training group, gray wafers in 2a of Figure 1, is then broken down into points used for training and those used for cross-validation.

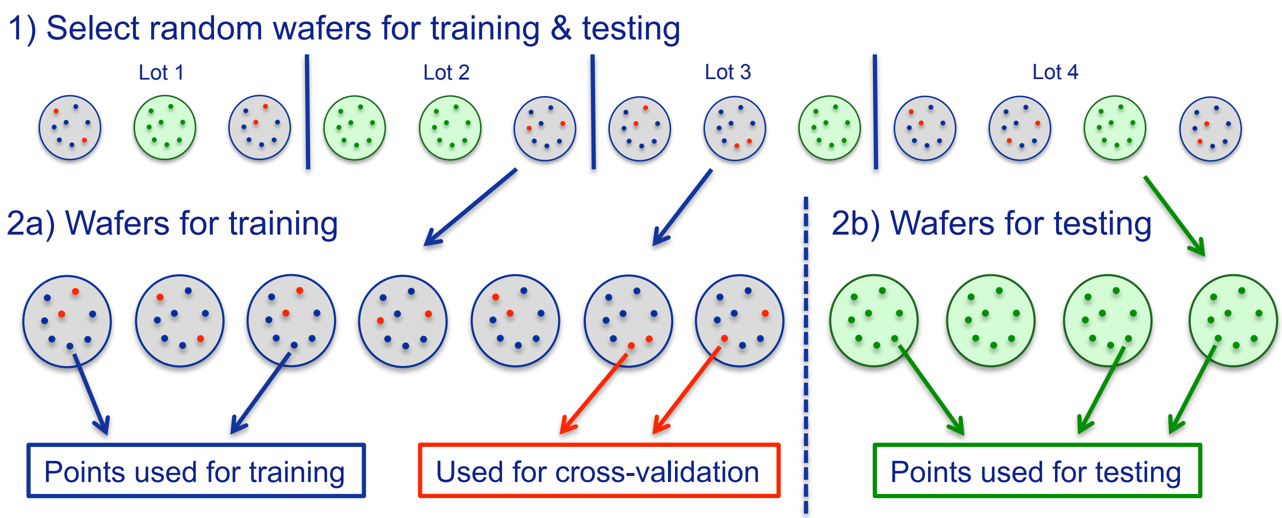

Figure 1: Data as broken down for training (gray wafers) and testing (green wafers) a neural network approximated function

2.2 Training the neural network for function approximation

To reduce the likelihood of overfitting the training dataset David MacKey's Bayesian framework [2] with regularization is employed. The regularization parameter for which is automatically optimized using the Gauss-Newton approximation to the Hessian matrix [3]. The process repeats in cycles until convergence, which is when the sum-squared error, the sum squared weights, and the effective number of parameters reach a constant value or till the cycle limit is reached. If the cycle limit is reached before convergence a new random sample of points is selected as training and cross-validation points within the training dataset (see 2a of Figure 1), while starting iterations toward convergence where the previous cycle stopped.

2.3 Testing the trained function

In Figure 2 we show how the trained function works. The function has three inputs and one output. The first input is wafer alignment metrology from all colors measured by the SMASH alignment sensor [4]. This includes wafer quality and residual values. The second is wafer-leveling metrology and the third is TWINSCAN context, wafer stage number along with field and target position per point on the wafer. The output of the function is a predicted overlay vector map. Because the function is already trained there is no excessive computational time to compute the wafer maps. The final result is a full lot of predicted overlay wafer maps. This includes wafers measured with integrated YieldStar metrology and those that were not.

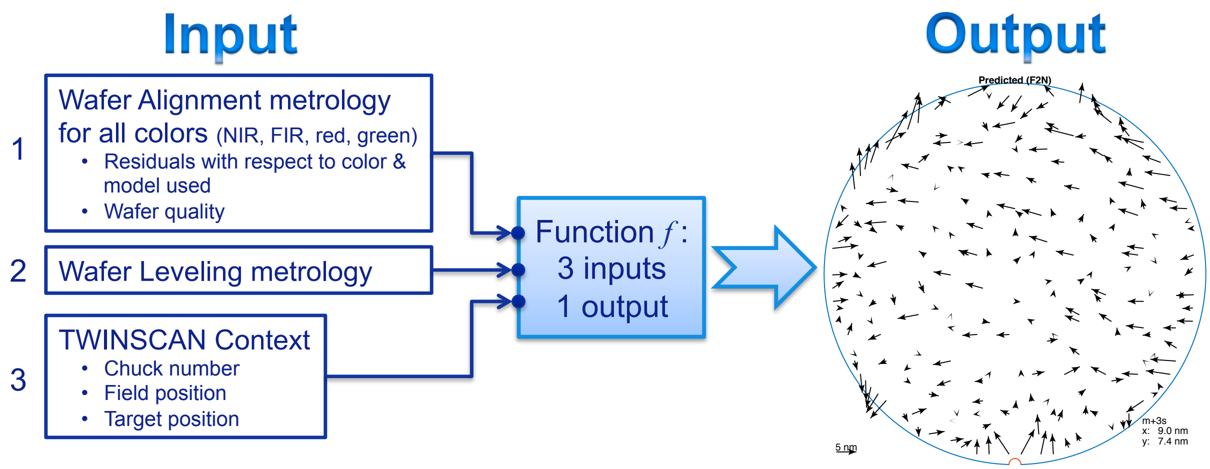

Figure 2: Flow diagram showing how the trained function works

2.4 Error between the measured and predicted overlay

Regression plots in Figure 3 show the wafer point-to-point correlation between the functions predicted output in Figure 2 (Output on Y axis in Figure 3) vs. YieldStar metrology (Target on X axis in Figure 3) for the wafers in both the training and testing groups (see 2a and 2b in Figure 1). Because the testing group of wafers in 2b of Figure 1 is "blind" to the training process we can use the testing wafer R-values for Overlay X and Overlay Y in Figure 3 to judge the performance of the trained function. In our case the R-values for both Overlay X and Y of the testing wafers are very close to the R-values of the training wafers. From this we conclude that with the automated regularization algorithm (section 2.2) the network generalizes well. That is we were successful in fitting only the contributors from the input that could be reliably correlated to the overlay metrology in the training group. It is important to note that residuals are unavoidable. They come from wafer noise, quality and the type of input metrology used. It is possible to obtain R-values closer to 1 with the wafers in the training group 2a of Figure 1, however this will have an inverse effect on the R-values of the testing wafers in 2b of Figure 1. The net result would be a degraded reliability of the approximated function due to overfitting.

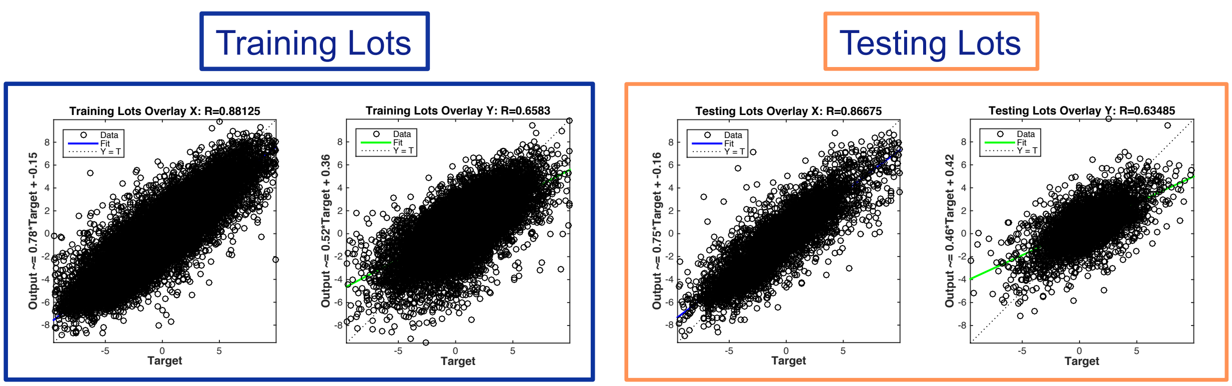

Figure 3: Regression plots of all wafers with predicted overlay outputs vs. YieldStar metrology

In Figure 4 we plot the data from Figure 3 as an overlay vector map. Figure 4a shows the average error between measured and predicted overlay with wafers from the training group. Figure 4b shows the average error between the measured and predicted overlay with wafers from the testing group. Figure 4c is the point-to-point difference between Figures 4a and 4b. With the scale of all three vector maps set to 1nm we see that the noise between the measured and predicted overlay is relatively consistent for both Training and testing groups of wafers. To understand the effect this has on the trained functions capability we cannot linearly add the wafer mean 3 sigma of the vector map in Figure 4 to the mean 3 sigma of the functions output, as shown in Figure 2. Instead we have to consider the error in Figure 4 as a plus or minus contribution per wafer coordinate location to the same location of any prediction from the trained function.

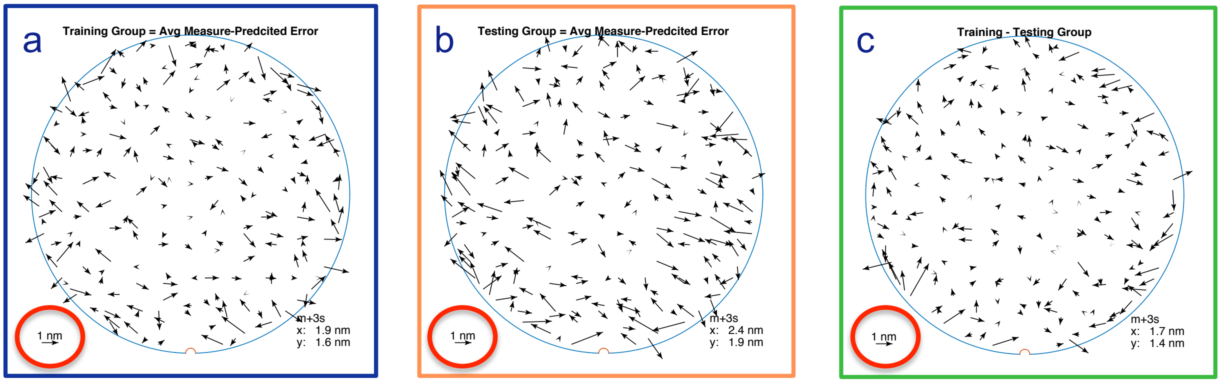

Figure 4: Average error between measured and predicted overlay for training and testing groups

3. RESULTS WITH PRODUCTION LOTS

SK hynix kindly provided the on product overlay data for our proof book analysis. The data for the 20nm DRAM layer we evaluated was collected over a 5-month period as it was prepared for high volume production by the integration team. During this time several processing steps were adjusted. Figure 5 shows mean 3 sigma (m3s) of Overlay X per wafer as three sequential plots. The top plot is the measured Overlay X per wafer, the middle plot is the predicted Overlay X per wafer and the bottom plot is the residual Overlay X per wafer. In Figure 6 we select the same wafer as it resides in each plot of Figure 4 to help visualize that the m3s were computed from full wafer vector maps. Wafer "a" of Figure 6 is the same wafer circled in the top plot of Figure 5. This wafer was measured with integrated YieldStar metrology. Wafer "b" of Figure 6 is the predicted vector map output from our approximated function (see Figure 2). Wafer "c" is the point-to-point delta between measured wafer "a" and predicted wafer "b", labeled as Residual.

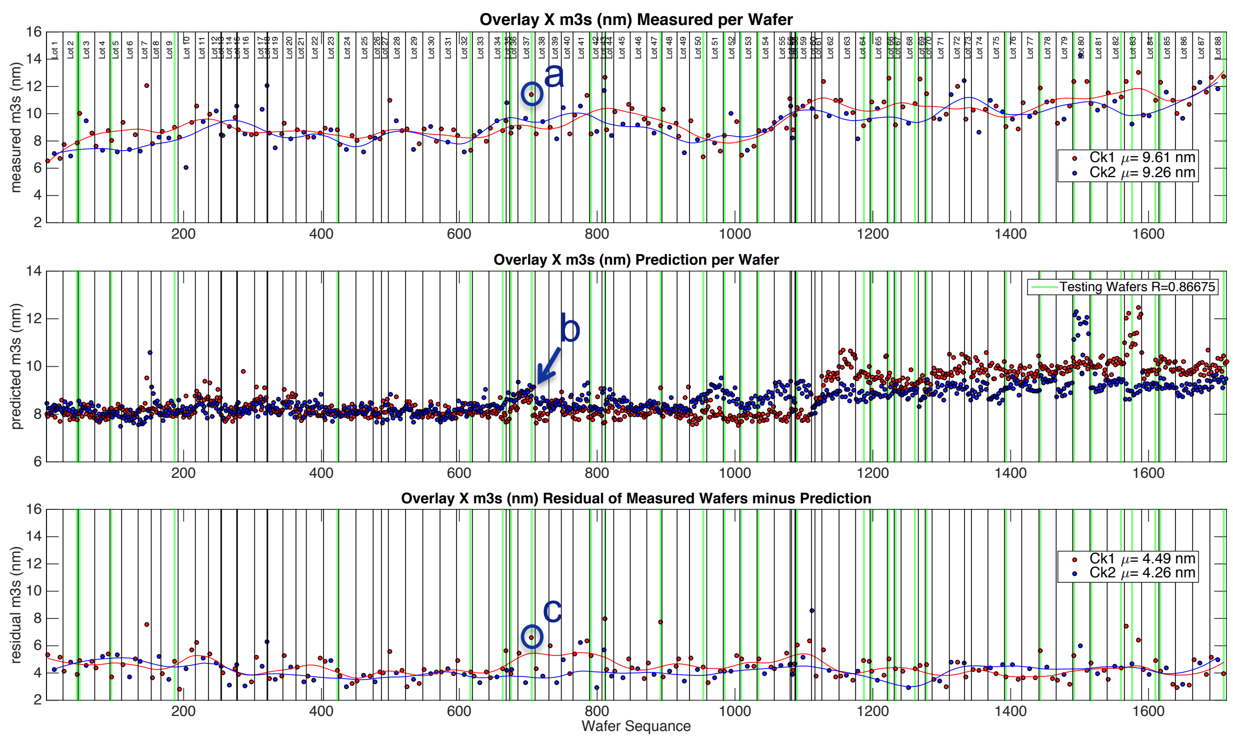

Figure 5: Mean 3 sigma (m3s) per wafer for the measured, predicted & residual Overlay X.

The legend of the middle plot in Figure 5 is a light green line that designates wafers from the Testing group (2a of Figure 1). These are the same wafers shown in Figure 3 as Testing Lots for Overlay X with an R-value of 0.86675, which gives a R^2 of 0.7513. Because these wafers are blind to the training step we use them to judge the performance of our approximated function. For Overlay X we can say that roughly 75% of the input parameters in Figure 2 could reliably be correlated to the overlay metrology.

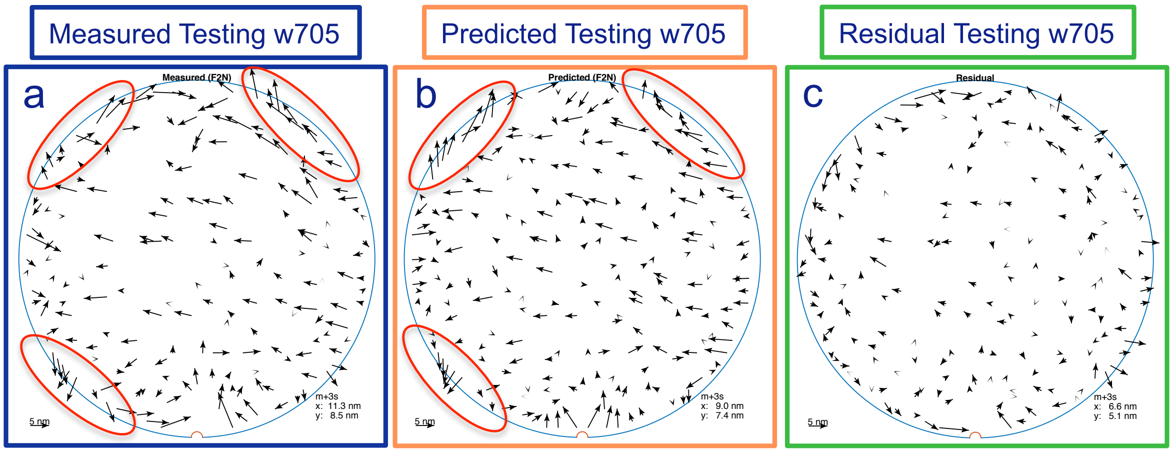

Figure 6: Single wafer vector map from each of the three plots in Figure 5

As Figure 5 shows Overlay X, Overlay Y for the same set of wafers is shown in Figure 7 as three sequential plots. The top plot shows measured overlay per wafer, the middle plot shows predicted overlay m3s and the bottom plot shows the residual between measured and predicted. In Figure 8 we select wafer 1,278 as it resides in each plot of Figure 7 to help visualize that m3s were computed from full wafer vector maps. Wafer "d" of Figure 8 is the same wafer circled in the top plot of Figure 7. Wafer "e" of Figure 8 is the predicted vector map output from our approximated function (as shown in Figure 2). Wafer "f" is the point-to-point delta between measured wafer "d" and predicted wafer "e", labeled as Residual.

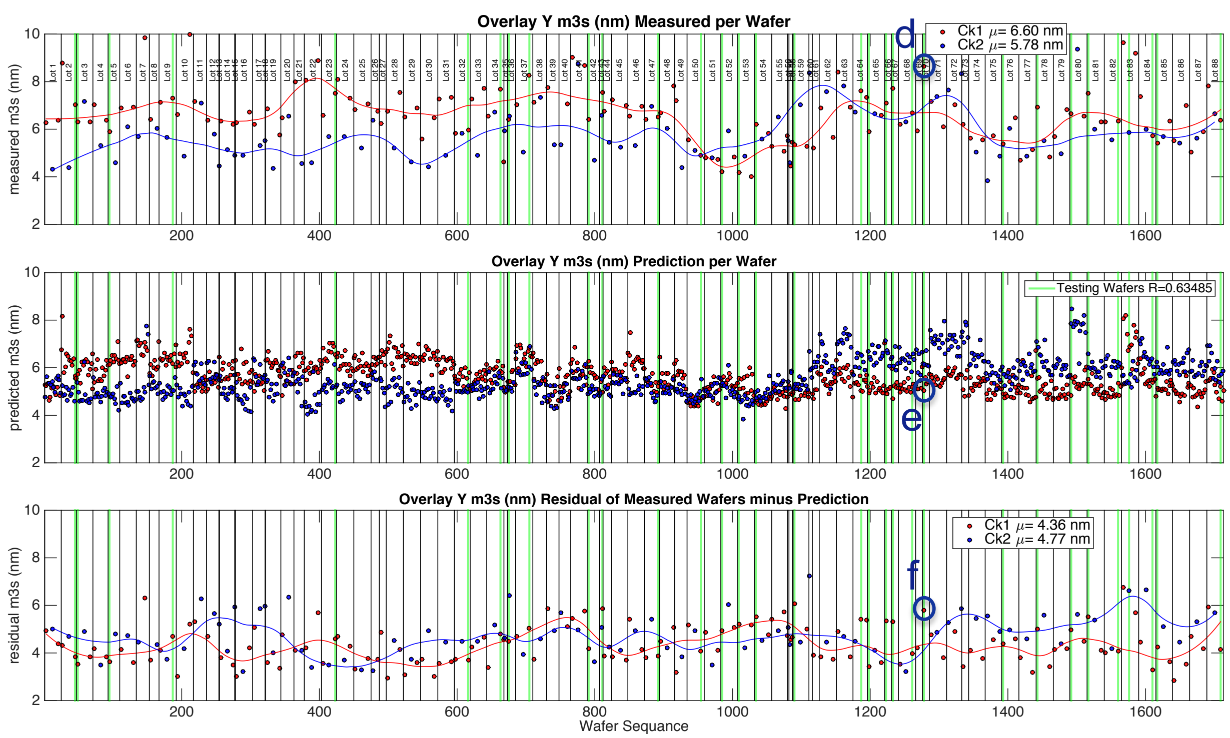

Figure 7: Mean 3 sigma per wafer for the measured, predicted & residual Overlay Y

In Figures 6 and 8 the predicted overlay vector maps closely resemble the measured overlay maps. Specifically take note of the region highlighted with red circles. It is these behaviors that we look to capture in the middle plots of Figures 5 and 7. Because we have a predicted wafer map for every wafer in the lot we can more easily see effects such as the delta between chuck 1 (red dots) and chuck 2 (blue dots).

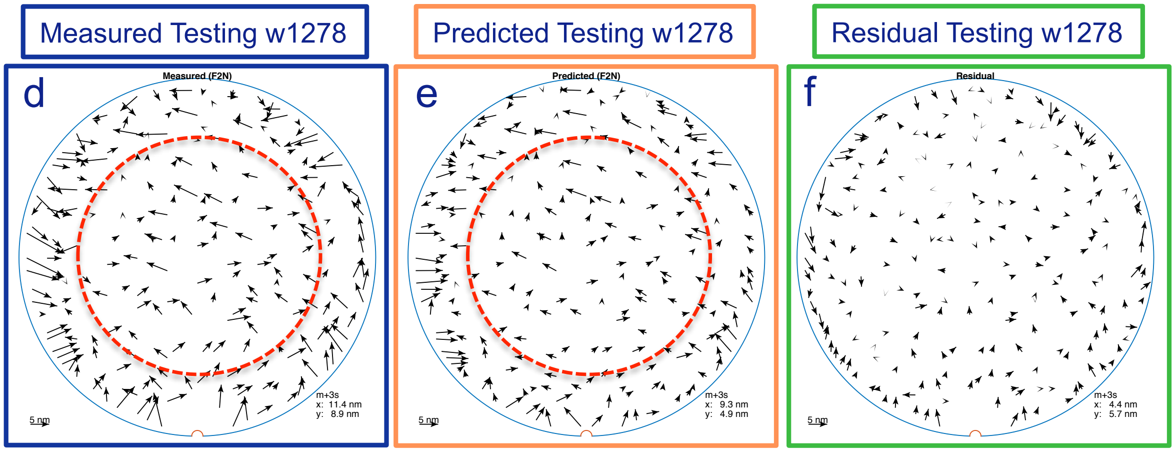

Figure 8: Single wafer vector map from each of the three plots in Figure 7

4. SUMMARY

With a machine learning technique known as function approximation, we were able to reliably train a function to predict the behavior of wafers we didn't measure with integrated metrology. When applied to data collected over a 5-month period of time our 220 measured wafers resulted in a prediction for 1,713 wafers. With the predicted overlay signature per wafer we can easily identify jumps in the overlay data where process was intentionally manipulated, facilitating process optimization. Looking into the point-to-point residual between the measured and predicted overlay we can flag wafers measured inline that exhibit an unexpected behavior. That is something other then the inputs we trained with is effecting the overlay signature. This can be used to remove a wafer from APC or to trigger an investigation. Moving forward, future work on this subject is open to any users with interest in exploring this application.

5. ACKNOWLEDGEMENTS

The authors would like to thank Jan Mulkens, Marcel Beems, Wolfgang Henke, Henry Megens, Peter ten Berge, Paul Luehrmann, Dick Verkleij, Frank van de Mast and Christophe Fouquet for their much valued feedback and support.

REFERENCES

[1] Schmitt-Weaver, E., Kubis, M., Henke, W., Slotboom, D., Hoogenboom, T., Mulkens, J., Coogans, M., ten Berge, P., Verkleij, D., van de Mast, F., "Overlay improvements using a real time machine learning algorithm" Proc. SPIE 9050, Metrology, Inspection, and Process Control for Microlithography XXVIII, 90501S (April 02, 2014); doi:10.1117/12.2046914; http://dx.doi.org/10.1117/12.2046914

[2] MacKay, D.J.C., "Bayesian interpolation," Neural Computation, Vol. 4, No. 3, 415–447 (May 1992). doi:10.1162/neco.1992.4.3.415; http://dx.doi.org/10.1162/neco.1992.4.3.415

[3] Foresee, F.D., and M.T. Hagan, "Gauss-Newton approximation to Bayesian regularization," Proceedings of the 1997 International Joint Conference on Neural Networks, 1930–1935 (1997).

[4] Hinnen, P., Depre, J., Tanaka, S., Lim, S.-Y., Brioso, O., et al, "Integration of a new alignment sensor for advanced technology nodes," Proc. SPIE 6520, Optical Microlithography XX, 652023 (March 26, 2007); doi:10.1117/12.712084; http://dx.doi.org/10.1117/12.712084