



HAL
open science

Prediction of Wafer Map Categories Using Wafer Acceptance Test Parameters in Semiconductor Manufacturing

Martin Lim, Anurag Sharma, Cheng Siong Chin, Tommy Yip, Jonathan Ong

► **To cite this version:**

Martin Lim, Anurag Sharma, Cheng Siong Chin, Tommy Yip, Jonathan Ong. Prediction of Wafer Map Categories Using Wafer Acceptance Test Parameters in Semiconductor Manufacturing. 18th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Jun 2022, Hersonissos, Greece. pp.136-144, 10.1007/978-3-031-08337-2_12 . hal-04668673

HAL Id: hal-04668673

<https://inria.hal.science/hal-04668673v1>

Submitted on 7 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Prediction of Wafer Map Categories using Wafer Acceptance Test Parameters in Semiconductor Manufacturing

Martin Lim Ying Song^{1,2}, Anurag Sharma², Cheng Siong Chin², Tommy Yip Chun Ming¹, Jonathan Ong Yoong Seang¹,

¹ Globalfoundries, Singapore 738406
martinyingsong.lim@globalfoundries.com

² NewRIIS, Singapore 609607

Abstract. The semiconductor industry is always looking for new solutions to maximize yield. Recently, the focus has been on utilizing the manufacturing data to help improve operational efficiency and early detection. This paper proposes a framework to find the best combination of machine learning models and data-balancing methods to predict specific wafer map signatures using Wafer Acceptance Test (WAT). WAT is a measurement test performed at multiple locations to identify poorly manufactured wafers. However, there were instances where wafers passed every measurement test but were found to have low yield. The proposed framework will be tested on real manufacturing data to demonstrate the viability of predicting wafer map signatures.

Keywords: Semiconductor Manufacturing · Machine Learning · Wafer Acceptance Test

1 Introduction

As global demand for Integrated Circuit (IC) chips is increasing, semiconductor manufacturers (fab) are under pressure to improve their yield and ensure that manufacturing errors are kept to a minimum. Wafers are the material in which multiple dies are manufactured on and every die contains the circuits found in IC chips. It is crucial for semiconductor manufacturing to ensure that all wafers meet manufacturing expectations. Wafer Acceptance Test (WAT) is used by fab engineers to identify poorly manufactured wafers and determine the number of dies that are usable (yield). After passing the WAT the wafers will go for Wafer Sort (WS) testing where individual die functionality is tested. A wafer map (WM) is then produced using the WS data that indicates which die passed or failed. The WM will be sent to the fab to evaluate the yield of each wafer and look for any process improvement to perform. One of the key things fab engineers look at when performing low-yield investigation is the wafer map signature as seen in Fig 1., which shows common low yield patterns. Identifying the WM signature is vital as it can identify possible manufacturing process failures. This is a time-

consuming process and has generated research interest in using machine learning [1 – 2] to automatically classify the signatures. However, there is a significant time gap (1 – 2 months) before the fab engineers get the WM from WS. Therefore, being able to

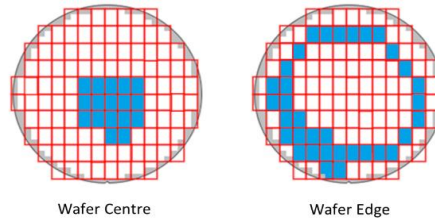


Fig. 1. Two of the most common wafer map signatures.

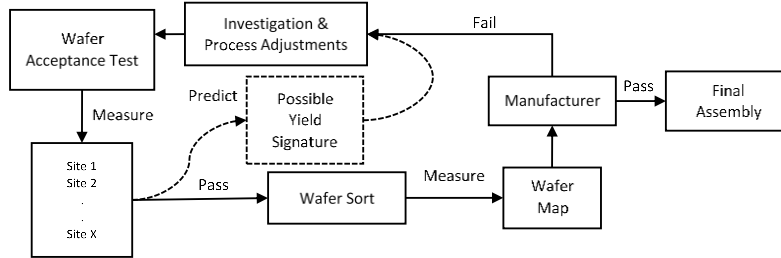


Fig. 2. Manufacturing flow after wafer fabrication stage.

predict the likelihood of the WM signature based on the yield predicted at different sections of the wafer as seen in Fig 2. will allow faster identification of possible manufacturing process error.

In the existing literature, there has already been research on using WAT measurements to predict wafer yield. This often involves using expert inputs and feature selection methods to identify key WAT measurements that explain the predicted yield value. Chien et al. [3] made use of experts to select relevant WAT measurements to be used in modeling. The experts chose the 12 measurements that was used in a modified Partial Least Square (mPLS) model. The model aims to prove the importance of the 12 measurements chosen by adding them into the model one at a time. There is a relative increase in R squared (R^2) score with every measure added. This proves that inputs from experts can help in improving prediction accuracy. However, the improvement gained from individual measurement is relatively low, and expert inputs can be biased to experience. The relative impact of other measures should also be considered that could be missed out from expert input. Using just expert input will not give new insights into possible manufacturing issues. Chan et al. [4] use a wide combination of feature and oversampling techniques to classify every die on the wafer to either pass or fail WS test. This is a more precise prediction as it is possible to generate the actual wafer map and classify using a wafer map classification model. However, the proposed framework

was only tested on two wafers from each lot which might be too small of a sample size to measure the predictive capabilities of this method. Jiang et al. [5] used WAT measurements together with descriptive data and used a one-hot encoder to create new parameters to represent the descriptive data. The data is then used to predict the yield at a final assembly where dies would have passed many tests prior, as seen in Fig 2. The data used here is a lot wider and it is predicting past other checks that were put in place to predict just the yield at final assembly. It does not particularly explain how all the tests up to the final assembly did not catch the predicted die failures earlier. Further studies would be required to understand this in greater detail. Xu et al. [6] made use of mutual information (MI) to find the correlation between WAT measurements and filtered the measurements using minimum Redundancy Maximum Relevance (mRMR). A deep belief network (DBN) was used to test the reliability of the filtering performed on the WAT measurements and genetic algorithm (GA) is used to rank the input features that were used in the DBN model. It was tested and it has achieved high accuracy score showing that using only feature selection is sufficient to improve prediction results. Further testing on using feature selection as the only approach to achieve high prediction score will be required to test the effectiveness. It is noted that proper reduction of WAT parameters is crucial for fab engineers to narrow down the problem when low yield occurs.

All previous works mentioned only focus on accurately predicting the actual yield by shrinking the WAT parameters for model training. However, unless the model can tell fab engineers which parameters to investigate when a low-yield wafer is predicted, it will still take significant time to identify the root cause manually. A different yield prediction approach that can identify WM signatures will allow fab engineers to identify the manufacturing process linked to each signature type. This paper proposes a best



Fig. 3. Framework to predict different wafer map classes.

combination of machine learning model and data-balancing methods to predict two major WM signatures, as seen in Fig 3.

The remainder of the paper is organized as follows: Section 2 introduces the methodology that will be used to predict the respective sections of the wafer. Section 3 explains the experimental setup used on real-life manufacturing data. Section 4 concludes the findings obtained from the experiment.

2 Methodology

This section will describe how site-level yield prediction will be performed based on site-level WAT measurement data. A combination of oversampling technique and machine learning algorithm will be used to find the best combination of models to predict for the two wafer map categories (Wafer Centre & Wafer Edge) as seen in Fig 1. These two categories are chosen as it has a high impact in terms of yield loss [1 – 2] and it covers majority of the wafer. Other categories will require additional data from a wafer map classification model to accurately identify each category which will not be discussed in this paper.

2.1 Data Preprocessing

2.1.1 WAT Measurement Site

To predict Wafer Map signatures would require a different type of dataset that includes the same WAT measurements that is performed at multiple locations (sites) throughout every wafer. The dataset for each signature will only include the measurement sites that overlaps with the signature that are identified earlier in Fig 1. The amount of measurement sites in each dataset will not be mentioned as it is confidential information. The location of the measurement sites across every wafer in a product will be the same. The chosen signatures (Wafer Center and Wafer Edge) are distinct from each other and there will be no overlapping sites in their respective dataset. The Wafer Edge signature will have significantly more measurements site as manufacturing processes are harder to control [8].

2.1.2 Wafer Yield Categories

The output used in this dataset will be the die yield that shows the number of dies that have passed all WS test. As mentioned earlier since there will be multiple sites on each wafer, the die yield will be calculated based on the Euclidean distance of the surrounding dies relative to where the location of the measurement for every wafer. As this is fundamentally a regression problem, the predicted values will show how many dies is likely to pass the WS test at every site. However, as the total number of dies at each site can be different making it difficult to determine the severity of the yield loss at each location.

To solve this, the regression output needs to have its own distinct categories to classify the severity of the yield loss at a given site. To do this a typical k-means clustering method is used to determine the threshold of yield loss at each site. There will be three category (Low, Medium & High) which will denote the amount yield loss relative to the total amount of dies associated with the site. Where the low category will represent the highest yield loss and vice versa. This will allow better distinction of how severe the yield loss at each site regardless of how many dies are associated with it.

2.2 Oversampling Algorithm

Class imbalance is a common problem when working with manufacturing data, as there is always a significantly larger quantity of good samples than bad samples. This results in most machine learning algorithms overfitting the majority class. To handle this, two over-sampling techniques are implemented to the training dataset to have the machine learning models be more sensitive to the minority class.

2.2.1 Synthetic Minority Over-sampling Technique (SMOTE)

This is a popular oversampling method [9] that uses k-NN to create synthetic samples from the minority data by computing the distance between the vectors in the space and multiplies a random value between zero to one. It is then added back to the original vector to be a new feature. The minority data will be oversampled to the maximum value from the majority data.

2.2.2 Adaptive Synthetic Sampling (ADASYN)

It uses an improved version of the SMOTE algorithm [10]. The key difference between the two is that ADASYN will calculate the number of synthetic minority class samples based on the amount of majority class samples that is within the minority class sample space.

2.3 Machine Learning Algorithm

2.3.1 eXtreme Gradient Boost (XGB)

XGB [11] is a machine learning algorithm that has high performance and accurate ensemble model and was made as an improvement over the traditional Gradient Boost Decision Tree (GBDT) where new trees are generated to add its loss function to the previous tree. XGB main. It also with missing values which is commonly found in any manufacturing environment due to equipment failure and human error. XGB objective function is defined as:

$$obj = \sum_{i=1}^n [l(y_i, \hat{y}_i)] + \sum_k \Omega(f_k) \quad (1)$$

where $\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2$ and l defines the loss function that measures the difference between the target value y_i and the predicted value \hat{y}_i . The regularization term is defined as $\Omega(f_k)$ where it is used to reduce the model complexity to avoid overfitting. It achieves this by considering the number of leaves (γ) and the number of leaf nodes (T) so that the best model can be achieved with the least resources possible.

2.3.2 Support Vector Regression (SVR)

SVR [12] is effective when there are no missing data, and there is no clarity on which parameters are deemed unnecessary for the model. It has less over-fitting when data re-sampling techniques are involved as it is designed to place a new sample into the nearest class 'label'. The only disadvantage it has is the large amount of training time required when running on a high dimension dataset.

3 Experimental Studies

3.1 Experiment Setup

The experiment is conducted on an existing wafer product and all manufacturing data is provided by Globalfoundries. The dataset consists of over 3000 wafers containing all measurement information on all site with no missing values. During final WAT measurement there is a total of 89 parameters tested on each site. The measurement sites will be divided into two Wafer Map signature (Wafer Center and Wafer Edge) as explained in Section 2.1. For each Wafer Map signature model, 70% of the site level WAT test data and its corresponding yield value will be used as the training set for the supervised regression model, and the remainder 30% of the data will be used as the testing set for test verification.

3.2 Model Training

This subsection will introduce how the training is conducted for the experiments, as shown in Fig 3. There will be a combination of two different class imbalance methods (ADASYN & SMOTE), and a combination of machine learning models (XGB & SVR) will be used to find the best combination of oversampling method and machine learning for the selected wafer map categories (Wafer Centre & Wafer Edge). The hyper parameters will be chosen by using K-fold cross-validation, where $K = 5$ and grid search to find the best sets of parameters and avoid over-fitting.

R-Squared (R^2) and Root Mean Square Error (RMSE) will be metrics used to determine how accurate the actual values predicted is. The closer the value for R^2 is to 1 the better the variance for all predicted results. However, a high R^2 will not be able to explain the actual problem on most manufacturing applications.

Therefore, a multi-class confusion matrix will indicate the category the predicted value will be classified under from the output clustering performed using k-means. The category will be classified from low yield – high yield (1 – 3). The confusion matrix will show the classification accuracy of each category with respect to the total quantity that was in each category.

3.3 Experiment Results

Both datasets for each of the Wafer Map Signature will be conducted with the experimental setup described above and the regression results can be seen in Table 1 as shown below.

Table 1. Comparison table between oversampling methods.

Regressor	Over-sampling Method	R ² (Edge)	R ² (Center)
XGB	None	-0.05	-0.01
	SMOTE	0.78	0.83
	ADASYN	0.78	0.8
SVR	None	-0.01	-0.37
	SMOTE	0.72	0.8
	ADASYN	0.76	0.75

To compare the performance between both machine learning models used in this experiment, XGB have a higher R² for both models across the board with 5% higher on average of the oversampling method used. This is likely due to the oversampling performed where a decision-tree based algorithm will perform better than a cluster-based algorithm. Choosing the correct oversampling method is also important and looking at the difference in results both methods give similar accuracy. Looking at the center signature model SMOTE have a higher R² with an accuracy score of 0.83 and 0.8 as compared to ADASYN with an accuracy score of 0.8 and 0.75 using XGB and SVR respectively. In the edge signature prediction, both SMOTE and ADASYN have similar performances using XGB with the same R² of 0.78 while ADASYN have a higher accuracy with the SVR model with an R² of 0.76 as compared to SMOTE with an R² of 0.72. This shows that SMOTE would on perform better in most scenarios as compared to ADASYN. While having no sampling will result in extremely poor R² values where it is unable to predict anywhere close to the target yield value.

However, R² only shows how accurate the model predicts relative to the target yield of each site. As explained in Section 2.1, the number of dies associated with each measurement site is different. Therefore, to better represent the yield loss at each site, the predicted value will be classified into three different yield categories (Low, Medium, High) yield. To see prediction accuracy in each category, a multi-class confusion matrix is used to demonstrate how well the transformed predicted values is classified as seen in Table 2. Where the predicted output will be transformed and classified into the three categories determined by the original yield output for each site. It also allows to see how far apart the prediction values will be.

Table 2. Number of correct classifications between oversampling methods.

Regressor	Over-sampling Method	Yield Category	Low (Edge)	Medium (Edge)	High (Edge)	Low (Center)	Medium (Center)	High (Center)
XGB	None	Low	15.46%	84.54%	0%	25.14%	74.86%	0%
		Medium	0%	100%	0%	0%	72.93%	27.07%
		High	0%	37.53%	62.47%	0%	0%	100%
	SMOTE	Low	68.42%	31.58%	0%	70.22%	29.78%	0%
		Medium	0%	100%	0%	0%	78.86%	21.14%
		High	0%	31.61%	68.39%	0%	0%	100%
	ADASYN	Low	77.71%	22.29%	0%	82.12%	17.88%	0%
		Medium	0%	100%	0%	0%	72.31%	27.69%
		High	0%	37.02%	62.98%	0%	0%	100%
SVR	None	Low	1.12%	85.73%	13.15%	7.32%	66.16%	26.52%
		Medium	0.38%	78.71%	20.91%	2.38%	61.25%	35.37%
		High	0.19%	76.16%	23.65%	2.23%	57.28%	40.49%
	SMOTE	Low	76.53%	22.45%	1.02%	80.5%	17.59%	0.91%
		Medium	12.85%	73.1%	14.05%	5.12%	71.95%	22.93%
		High	6.4%	72.8%	20.8%	5.22%	52.58%	42.2%
	ADASYN	Low	68.79%	25.21%	6%	73.44%	17.57%	8.99%
		Medium	3.42%	73.5%	23.08%	4.54%	61.95%	33.51%
		High	1.41%	71.51%	27.08%	2.08%	57.98%	39.94%

Looking at Table 2, it is clearer in explaining how each of the R^2 score came to be. As mentioned earlier, using the original dataset with no oversampling will result in the model being unable to predict any of the target values by looking at just the R^2 . However, looking closely at Table 2, the prediction inaccuracy only occurs mostly at the Low Yield and High Yield category. This will clearly indicate where the class imbalance occurs for each dataset and a more precise oversampling to target specific categories could be performed. In this experiment, only the Low Yield category is

oversampled which can be seen in the increase in the classification rate by 83% on average in both XGB and SVR for both signatures while the classification rate for Medium Yield and High Yield category remains close to the unsampled dataset.

The best machine learning model and oversampling combination can also be determined using the results shown in Table 2 whereas looking at the R^2 values in Figure 1 would not result in a clear choice. XGB performs better with ADASYN as compared to SMOTE with a classification score of 77.71% and 82.21% with ADASYN and a classification score of 68.42% and 70.22% for Wafer Edge and Wafer Center categories respectively. While SVR performs better with SMOTE with a classification score of 76.53% and 80.5% with SMOTE and a classification score of 68.79% and 73.44% for Wafer Edge and Wafer Center categories respectively.

However, XGB is the better model of the two as there are no cases in which an incorrectly predicted yield category crosses to the extreme end (i.e., Low Yield to High Yield) and is only one category apart from its intended category making it less prone to severe False Positive and False Negative classification results.

4 Conclusions

The ability to forecast potential die yield should be significant in helping fab engineers be aware of potential process issues during manufacturing earlier. This paper proposes a framework that uses WAT measurements to identify potential wafer map signature, and to compare the impact of oversampling between different types of machine learning models. It makes use of the WAT site measurement that can be used to correlate to the other kind of wafer map signatures. This is an early concept that requires more accurate data from Wafer Map classification and a better understanding of how WAT can potentially be used to predict wafer yield.

5 Acknowledgement

I would like to thank Globalfoundries for providing data and support for the published work.

References

1. J. Shim, S. Kang, and S. Cho, "Active Learning of Convolutional Neural Network for Cost-Effective Wafer Map Pattern Classification," in *IEEE Transactions on Semiconductor Manufacturing*, vol. 33, no. 2, pp. 258-266, May 2020, doi: 10.1109/TSM.2020.2974867.
2. L. L. -Y. Chen *et al.*, "Semi-Supervised Framework for Wafer Defect Pattern Recognition with Enhanced Labeling," *2021 IEEE International Test Conference (ITC)*, 2021, pp. 208-212, doi: 10.1109/ITC50571.2021.00029.
3. C. -F. Chien, P. -C. Lee, R. Dou, Y. -J. Chen and C. -C. Chen, "Modeling collinear WATs for parametric yield enhancement in semiconductor manufacturing," *2017 13th IEEE Conference on Automation Science and Engineering (CASE)*, 2017, pp. 739-743, doi: 10.1109/COASE.2017.8256192.

4. X. Chen, Y. Zhao, H. Lü, X. Shao, C. Chen, and Y. Huang, "A Machine Learning-based Approach for Failure Prediction at Cell Level based on Wafer Acceptance Test Parameters," *2021 IEEE Microelectronics Design & Test Symposium (MDTS)*, 2021, pp. 1-5, doi: 10.1109/MDTS52103.2021.9476151.
5. D. Jiang, W. Lin, and N. Raghavan, "A Novel Framework for Semiconductor Manufacturing Final Test Yield Classification Using Machine Learning Techniques," in *IEEE Access*, vol. 8, pp. 197885-197895, 2020.
6. H. Xu, J. Zhang, Y. Lv and P. Zheng, "Hybrid Feature Selection for Wafer Acceptance Test Parameters in Semiconductor Manufacturing," in *IEEE Access*, vol. 8, pp. 17320-17330, 2020, doi: 10.1109/ACCESS.2020.2966520.
7. T. C. Tin *et al.*, "A Realizable Overlay Virtual Metrology System in Semiconductor Manufacturing: Proposal, Challenges and Future Perspective," in *IEEE Access*, vol. 9, pp. 65418-65439, 2021, doi: 10.1109/ACCESS.2021.3076193.
8. M. P. -L. Ooi, E. Kwang Joo Sim, Y. C. Kuang, L. Kleeman, C. Chan and S. Demidenko, "Automatic Defect Cluster Extraction for Semiconductor Wafers," *2010 IEEE Instrumentation & Measurement Technology Conference Proceedings*, 2010, pp. 1024-1029, doi: 10.1109/IMTC.2010.5488012.
9. Chawla, Nitesh V., et al. "SMOTE: synthetic minority over-sampling technique." *Journal of artificial intelligence research* 16 (2002): 321-357.
10. H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," in 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). IEEE, 2008, pp. 1322–1328.
11. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Chen, K. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4), 1-4.
12. Awad, Mariette, and Rahul Khanna. "Support vector regression." *Efficient learning machines*. Apress, Berkeley, CA, 2015. 67-80.