Towards Big Data Visualization for Monitoring and Diagnostics of High Volume Semiconductor Manufacturing

Dimitra Gkorou, Alexander Ypma, George Tsirogiannis, Manuel Giollo, Dag Sonntag, Geert Vinken, Richard van Haren, Robert Jan van Wijk, Jelle Nije, Tom Hoogenboom
ASML, De Run 6501, Veldhoven 5504DR, the Netherlands

ABSTRACT

In semiconductor manufacturing, continuous on-line monitoring prevents production stop and yield loss. The challenges towards this accomplishment are: 1) the complexity of lithography machines which are composed of hundreds of mechanical and optical components, 2) the high rate and volume data acquisition from different lithography and metrology machines, and 3) the scarcity of performance measurements due to their cost. This paper addresses these challenges by 1) visualizing and ranking the most relevant factors to a performance metric , 2) organizing efficiently Big Data from different sources and 3) predicting the performance with machine learning when measurements are lacking. Even though this project targets semiconductor manufacturing, its methodology is applicable to any case of monitoring complex systems, with many potentially interesting features, and imbalanced datasets.

KEYWORDS

continuous monitoring of high volume manufacturing, anomaly detection, analytics, visualization of high dimensional data, machine learning, data science

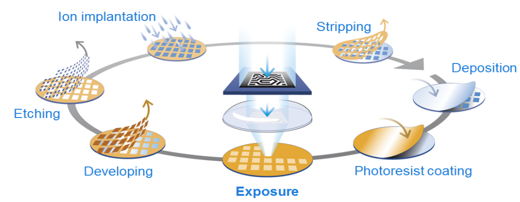
1 INTRODUCTION

Integrated Circuits (ICs) are defined on a *wafer*, , a round slice of semiconductor material, in layers whose number can easily be 30 or more, as we see in Figure 1a. *Patterning* is about optimally transferring a pattern from mask to wafer. Misalignment of the exposed layers results in defective chips due to electrical connection failures and short circuits. *Overlay error* is the misalignment between the layers printed on top of each other in nanometers and it is one of the most dominant factors for yield loss [7, 8] (Figure 1b). Any lithography machine is composed of hundreds of mechanical and optical parts whose malfunction may cause overlay error. Identifying and monitoring online the main factors contributing to overlay error and performance-predictive modeling enables early decision making against yield loss.

Nevertheless, lithography is just a part of the processing steps needed creating ICs. In a fab, a wafer passes multiple other process steps that could cause overlay error (Figure 1a). Process variations on top of the scanner performance impose significant challenges

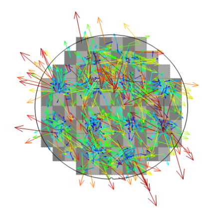
CF'17, Siena,Italy

DOI: http://dx.doi.org/10.1145/3075564.3078883



(a) Process steps of a wafer.

(b) Schematic picture of overlay.



(c) Overlay measurement between two layers.

Figure 1: Exposure is performed on lithography machines and overlay measures its accuracy.

for process engineers during ramp in order to get to acceptable yield levels. Customers are observing significant overlay during the manufacturing optimization for a new product due to suboptimal fab process. It takes several months for very expert fab engineers to fine tune tool recipes before getting to a stable production outcome. Building up the knowledge and experience to deal with such problems is becoming harder for engineers, due to the complexity related to extreme miniaturization. In order to support fab engineers on the optimization of overlay error, we need to identify the most relevant parameters to the observed overlay error, combine them in *overlay error indicators* that can be monitored through the exposed layers of a wafer, and visualize them in a meaningful way so that fab engineers can provide feedback to the selected parameters.

Big Data and Machine Learning transformed the landscape of high tech industry over the last few years, and they proved to be able to generate knowledge automatically by exploiting the available data. This work reports how these techniques can serve as a complementary tool to engineer knowledge in order to speed up diagnostics in fabs in terms of overlay.

Our methodology focuses on optimizing patterning in terms of overlay error which implies a co-optimization of several machine parameters and components. In this context advanced data analysis, modeling, and machine learning offer powerful tools to understand, monitor, and combine these parameters. Our methodology has been developed in the context of ASML Litho InSight (LIS) analytics platform. LIS analytics carries out data analysis, monitoring, trending,

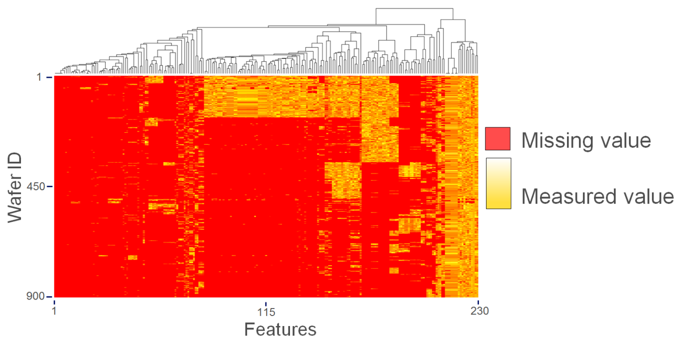


Figure 2: A dendrogram of the dataset. Each cell shows the value of features on a specific wafer.

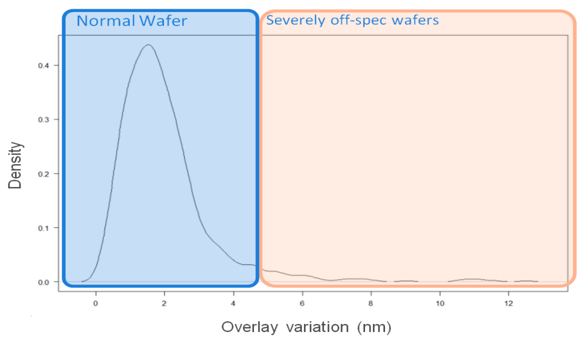


Figure 3: Probability Distribution Function of Overlay variation. A minority of wafers have severely off-spec overlay

and down drilling for the semiconductor industry. We developed our methodology around the following aspects:

(1) Which are the most important factors that contribute to overlay error, do they provide enough information to identify root causes, and how do they interact with each other?
(2) Which is the most effective way to visualize these factors in a structured way so that similarly exposed wafers are grouped together and that the resulting visualization enables monitoring and diagnosis?
(3) Can we estimate the overlay error of the unmeasured wafers?

2 DATA CHARACTERISTICS

We developed our methodology in order to support lithography experts at customers fabs to stabilize process variations causing overlay. We collected measurements of process and lithography steps as well as overlay measurement. In this section, we present the characteristics of fab and overlay measurements highlighting the challenges that they pose, for the Machine Learning solution. It should be noted that all the properties reported below are very typical in the semiconductor industry data. Therefore, it is important to develop methods and workflows that can deal effectively with these challenges in order to speed up the diagnosis process of fab problems.

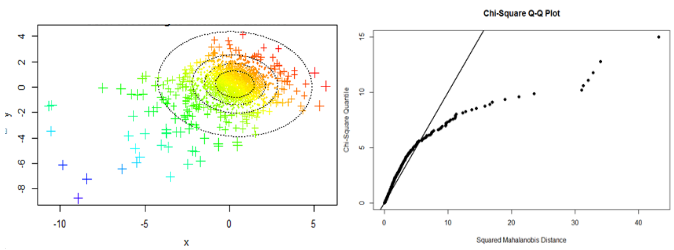
Missing Values Fab process measurements range from wafer layer data to fab equipment configuration like etching and Chemical Mechanical Polishing (CMP) time. As can be seen in Figure 2, most of this information is collected only on a small number of wafers, mostly due to time constraints in the production. On average, just 25% of the information is retrieved. Even more important, there are clearly patterns in the sampling of features. When few measures are gathered, then there is an incentive in making a good wafer characterization, which helps problem identification. The same approach is also used in the medical diagnosis process [6]. This *conditional sampling* of features shows an hierarchical structure, and can be represented with a dendrogram. In Figure 2, such a dendrogram is shown for a dataset of 900 wafers where 230 different type of fab information was measured.

Unbalanced datasets The customer goal, the maximization of yield, is typically achieved by minimizing the variability during production process. Product imperfections happen either 1) due to set up of a new product, which requires a new configuration or 2) due to malfunction of some fab machinery. As a result, samples have high chance to be very similar to each other, and it becomes difficult from the statistical point of view to learn about the impact of a feature variation. Our goal is the identification of which wafer features best explain why there are *normal wafers* and *off-spec wafers* (see Figure 3). These two categories represent 95% and 5% of the dataset, respectively. Dealing with *unbalanced datasets* with a large number of missing values is a hard challenge, which requires robust statistical techniques.

Curse of Dimensionality Feature redundancy is another big challenge in this dataset. A number of measures taken from different tools are actually looking at the same effect, therefore leading to *collinearity*. From the analytical point of view, this leads to issues in understanding the causality of effects and to spurious findings. In the semiconductor industry, the number of features p is greater than the number of samples n. For each wafer, about 20,000 measurements are being collected. The *curse of dimensionality* [3] and missing data will prevent us from building a model of overlay based on the measured information.

Partially labeled datasets Overlay measurements can be a bottleneck in production as it is time-consuming and a wafer is exposed several times. Thus, it is performed only to a small subset of wafers. On the other hand, lithography machines produce detailed reports with abundant information for all the wafer exposures. This fact results in *partially labeled* datasets which challenge analysis and performance-predictive modeling. We note that an ASML lithography machine exposes about one wafer every 15 sec.

Noisy data The overlay is measured with specialized metrology equipment after the exposure by using reference markers printed at each layer (Figure 1c). Physical deformation of the markers due to non litho processing may cause noise on the overlay measurements. Even small deformations can have significant contribution to the quality of measurement. As a result, even labeled data should be treated carefully before being used as input to machine learning algorithms and methods.



(a) The distribution of overlay measurements.

(b) Q-Q plot

Figure 4: Finding off-spec wafers.

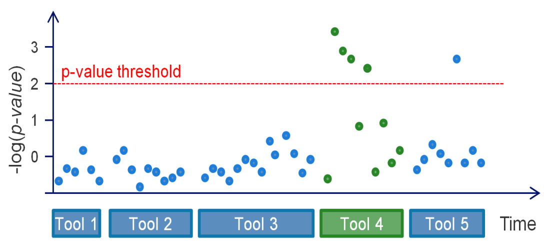


Figure 5: The result of association tests. Features beyond the significance threshold shows a link with respect to overlay variation.

3 METHODOLOGY

The support of fab engineers consists in linking the most relevant parameters to the observed overlay error, combining them in *overlay error indicators* that can be monitored through the exposed layers of a wafer, and visualizing them in a meaningful way so that fab engineers can act upon the selected parameters. An appropriate Machine Learning work-flow has to fulfill the problems requirements introduced by the data characteristics.

We perform *feature selection* methods on parameters measured during wafer exposure in lithography machines as well as additionally created features. To select the most relevant features, we rank them in terms of their ability to predict overlay error and we select the highest ranked features. To deal with the problems of missing values, unbalanced data, and the lack variability on the data samples, we consider a classification problem where the goal is to predict off-spec wafers with high overlay error. Thus, data preprocessing includes the characterization of off-spec wafes (outlier detection as also presented in Figure 4). Considering that fab mesurements contain the preselection of a relevant feature subset.The potentially relevant exposure parameters are either categorical or numerical, they exhibit interactions and they connect with overlay error in a non-linear way. Thus, we use a Random Forest [1]. Combining the most important features from wafer exposure, we develop *overlay error indicators* which are online monitored together with their interactions, as indicated in Figure 10. Fab engineers evaluated positively the proposed indicators. An illustration of this Machine Learning flow is presented in Figure 7.

3.1 Preprocessing

To characterize off-spec wafers, we perform *multivariate outlier detection* on overlay measurements using the ordered squared Mahalanobis distances of the overlay measurements from their mean, where the mean and the covariance matrix are estimated robustly. More precisely, we compare the computed squared Mahalanobis distances of the overlay measurements with the quantiles of the corresponding χ^2 distribution. By using the Quantile-Quantile plot, we graphically observe outliers (Figure 4b) as described in [4, 5].

Understanding the root cause of wafer quality variation requires to focus on feature *causality*. *Association tests* are used in these situations to narrow down the number of significant variables that should be kept into account during causality studies [6]. Depending on the variable distributions, *t-test*, χ^2*-test* and *Wilcoxon-Mann-Whitney test* can be used to generate a *p-value*. These tests use as input just a single feature and the response, therefore they minimize the impact of missing features. *Multiple testing correction* must also be taken into account to select a proper significance cutoff. It is interesting to note that a number of measures are obtained for a certain wafer after each tool process. Typically, multiple features show strong association with overlay change if they are obtained from a defective tool (see green tool in Figure 5).

3.2 Modeling

After data preprocessing, a smaller subset of relevant features is kept for further processing. In order to build an effective predictive model with non-linear terms and assess the feature importance, we use a *Random Forest* [1] which grows an ensemble of *decision trees*. The algorithm samples any input dataset and builds each decision tree using a limited set of features. This subsampling procedure helps to estimate tree performance on unseen examples and deals with noisy data. By repeating this process, a collection of trees can be used to predict overlay variation along with an average prediction performance. Starting from this model and its expected performance, one can also estimate variable importance. Given a feature f in the dataset, one can permute its value and evaluate once again the model effectiveness. The relevance measure for each feature is obtained by looking at the performance difference before and after dataset permutation [2]. This whole process can be repeated few times with *cross-validation*, producing confidence interval for both performance and variable importance (see Figure 6).

3.3 Finding Overlay Error Indicators

Combining the most important features from wafer exposure, we develop *overlay error indicators* which are online monitored together with their interactions, as indicated in Figure 10. The developed indicators can be computed for the unlabeled wafers as well, allowing a fab engineer to conclude the overlay error of the wafers that remained unmeasured for overlay. Such an indicator is presented in Figure 8 for reticle heating. This indicator was computed on a dataset for which we had overlay measurements for only 125 lots while the full dataset consisted of 13, 000 lots. The most important features depend on the *exposure recipe* of a product. Examples of relevant features include the reticle transmission factor, the reticle align magnification drift, the dose, the throughput, and the usage of the same reticle over consecutive exposures.

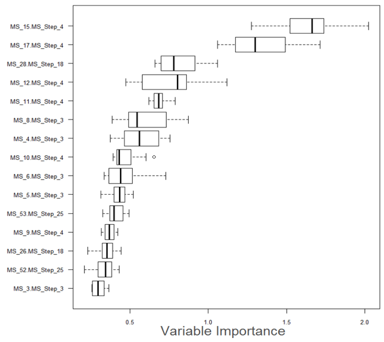


Figure 6: The importance of features in explaining overlay variability, as evaluated by a Random Forest algorithm. We obtain an estimate of the importance variability with cross validation.

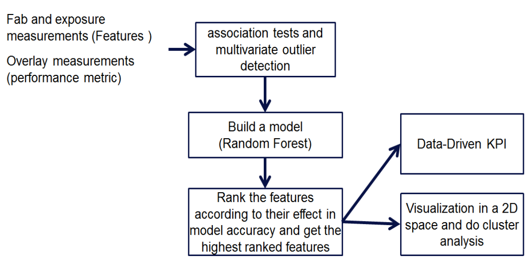


Figure 7: The workflow adopted to process data.

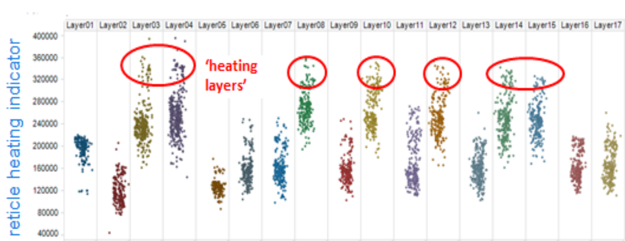


Figure 8: A reticle heating indicator observed through the consecutive exposure layers.

3.4 Visualization

Visualizing the most important features in a two dimensional space is essential for observing meaningful patterns and interconnections that are usually difficult to identify in high dimensional spaces. Such a visualization can actively drive user input to the algorithm. We note that visualization is currently an unsupervised step.

To visualize the most important features, we perform t-Distributed stochastic neighbor embedding (t-SNE)[9], a nonlinear dimensionality reduction technique for embedding high-dimensional data into a two dimensional space. In t-SNE, points that are similar in the high-dimensional space are drawn close in the two dimensional space while dissimilar points are drawn far apart. Figure 9 presents the two dimensional map of the eight most important measurements for

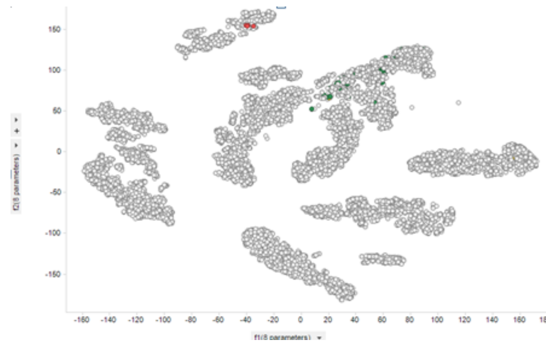


Figure 9: Groups of similarly exposed wafers.The white datapoints indicate unlabeled information, the red ones indicate high overlay error, and the green ones indicate low overlay error.

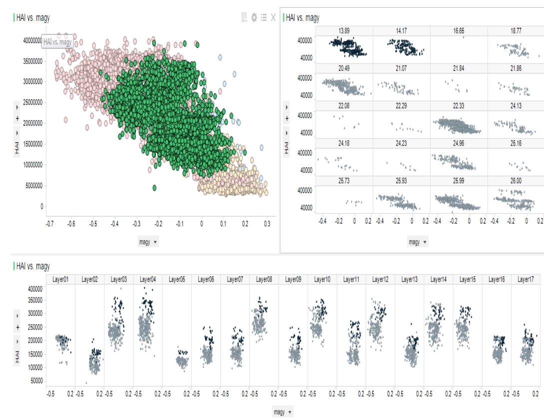


Figure 10: Online monitoring of overlay error indicators using exposure information.

reticle heating as computed by t-SNE. This embedding technique defines a probability distribution over pairs of high-dimensional points so that similar points have a high probability of being selected. Similarly, it defines a similar probability distribution over the points in the low-dimensional map. Finally, it minimizes the Kullback-Leibler divergence between the two distributions.

After the visualization in the 2D space, the user can obtain valuable information regarding the relation among the data points. Clusters in 2D space indicate similarities of the original data, properties and interconnections usually difficult to identify in high dimensional spaces. Particularly in Figure 9, we observe the 2D visualization of the reticle heating features described in the previous subsection. A user can observe the formed clusters. Using overlay metrology of the 125 exposures, we label with red the high heating exposures and with green the non-heating exposures and we observe that these types of lots belong to different clusters. One distinguishing factor between the cluster with heating lots and the cluster with the non-heating lots is image size in x direction.

3.5 Implications

We show that an effective map from exposure data to overlay measurements can be made, possibly enabling more efficient monitoring.

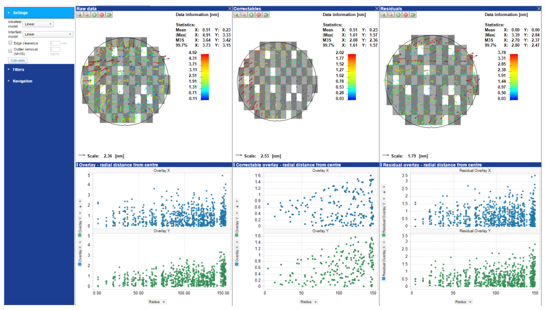


Figure 11: An example of a trend plot dashboard.

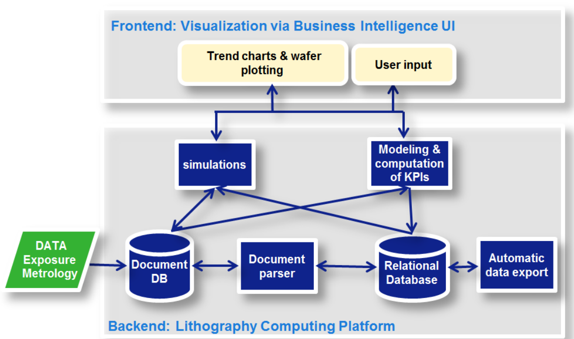


Figure 12: The architecture of Litho Insight Analytics & Visualization

For example, using overlay metrology to label the clustered data, we observe that exposures of high and low overlay belong to different clusters. The results of our analysis and visualization have been verified and adopted by fab engineers. The advantages of our approach are:

- Exposure-based performance indicators are easily monitored across the different layers of a wafer. New arriving datapoints are connected smoothly with already existent historical data and the indicators are updated automatically.
- Quick investigation of non-trivial patterns and gaps between exposure and metrology measurements from a domain expert without any machine learning background, which can guide future measurements. Based on the underlying exposure data, a domain expert without any machine learning background, can identify wafers in an early phase wafers that that are candidates for additional metrology.
- A user of our methodology can quickly respond to the observed patterns and take decision and actions.
- Possibly increased informativeness of overlay metrology efforts in the fab.
- Despite the focus of this demo on lithography, our approach is applicable to any other domain with imbalanced datasets, missing values, and many relevant features.

4 LITHO INSIGHT ANALYTICS & VISUALIZATION

The presented work will be part of Litho Insight (LIS) Analytics & Visualization. In the context of LIS, data from multiple lithography and metrology machines is combined efficiently and stored to a database system. Dashboards developed on top of this database, allow engineers to monitor on-product performance, diagnosing problems, drilling-down from top-level dashboards into detailed visualizations, and quickly find areas for improvements. Figure 12 shows a representation of the architecture of LIS Analytics & Visualization. This platform has been developed to optimize reliability and fast query retrieval.

The organized datasets are visualized and analyzed in a way that facilitates the comparison of different machines and the intuitive monitoring of performance indicators across the different exposed layers. Figures 10 and 11 show examples of implemented dashboards. In our dashboards, we adopt a *hierarchical visualization* scheme. Starting from a machine-level view (Figure 10), a user can zoom in to products, lots (units typically containing 25 wafers), and finally to a single wafer.

5 CONCLUSION

We focus on continuous monitoring of high volume manufacturing for a targeted performance metric. Data from multiple lithography and metrology machines is integrated, visualized and analyzed. The analysis contains feature selection methods, dimensionality reduction for advanced visualization, and regression-based prediction of overlay error. Our integrated analysis flow and visualizations help domain experts to fine-tune the production line and interpret patterns in heterogeneous data. Our visualizations are intuitive and applicable to other fields other than lithography.

REFERENCES

[1] Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (2001), 5–32. DOI: http://dx.doi.org/10.1023/A:1010933404324

[2] Houtao Deng and George Runger. 2012. Feature Selection via Regularized Trees. *The 2012 International Joint Conference on Neural Networks (IJCNN), IEEE* (2012).

[3] Pedro Domingos. 2012. A Few Useful Things to Know About Machine Learning. *Commun. ACM* 55, 10 (2012), 78–87.

[4] Peter Filzmoser. 2017. Multivariate Outlier Detection Based on Robust Methods. CRAN repository. (Jan. 2017). https://CRAN.R-project.org/package=mvoutlier

[5] Peter Filzmoser, Robert G. Garrett, and Clemens Reimann. 2005. Multivariate Outlier Detection in Exploration Geochemistry. *Comput. Geosci.* 31, 5 (June 2005), 579–587.

[6] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. 2002. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* 46, 1 (2002), 389–422. DOI: http://dx.doi.org/10.1023/A:1012487302797

[7] A. Lam, F. Pasqualini, J. de Caunes, and M. Gatefait. 2010. Overlay breakdown methodology on immersion scanner. *Proc. SPIE* 7638 (2010).

[8] A. Lam, A. Ypma, M. Gatefait, D. Deckers, A. Koopman, R. van Haren, and J Beltman. 2015. Pattern recognition and data mining techniques to identify factors in wafer processing and control determining overlay error. *Proc. SPIE* 9424 (2015).

[9] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9 (2008).

[10] M. Mitchell Waldrop. 2016. More than Moore. *Nature* 530 (2016), 144 âĂŞ 147.

[11] A. Ypma, J. Menger, D. Deckers, D. Han, A. Koopman, I. Lyulina, S. Middlebrooks, R. Haren Van, and J. Wildenberg. 2015. Methods and apparatus for obtaining diagnostic information relating to an industrial process. Patent WO2015049087 A1. (2015).