Real-Time Fault Detection in Semiconductor Using One-Class Support Vector Machines

Ali Hajj Hassan, Sophie Lambert-Lacroix, and Francois Pasqualini

Abstract—In this paper, we propose a real-time fault detection system for the semiconductor domain, which aims to detect abnormal wafers from a recent history of electrical measurements. It is based on a dynamic model which uses our filter method as feature selection approach, and one-class support vector machines algorithm for classification task. The dynamicity of the model is ensured by updating the database through a temporal moving window. Two scenarios for updating the moving window are proposed. In order to prove the efficiency of our system, we compare it to an alternative detection system based on the Hotelling's T^2 test. Experiments are conducted on two real-world semiconductor datasets. Results show that our system outperforms the alternative system, and can provide an efficient way for real-time fault detection.

Index Terms—Real-time detection, feature selection, one-class support vector machines, semiconductor.

I. INTRODUCTION

Nowadays, the control of manufacturing processes is an essential task to ensure consistently safe operation and high quality production. This is challenging particularly when processes have a large number of operations and complex systems, which is the case in manufacturing process of semiconductor devices and integrated circuits. Early and accurate detection of faults is then required for maintaining a process at its optimal condition, and reducing manufacturing costs.

Once the manufacturing process of semiconductor ends, an electrical test, called Parametric Test (PT), is performed. PT aims to detect within shortest possible time the abnormal wafers (semiconductor material used in manufacturing of semiconductor devices) by looking at a set of static electrical parameters measured on multiple sites of each wafer.

The purpose of this work is to implement an automatic real-time detection system at PT level. Based on a multivariate statistical approach, this system aims to detect abnormal wafers through a moving temporal window of electrical measurements.

Multivariate statistical approaches have been successfully used for monitoring industrial processes [1]–[3]. Principal Component Analysis (PCA) was considered to develop respectively a static (off-line testing) and dynamic (in-line testing) models for fault detection in biological Wastewater Treatment Plant (WWTP) [4], [5]. PCA was also considered in [3] to detect faults in a semiconductor etch process. PCA is one of the most widely multivariate techniques used for extracting relevant information from high dimensional data. The goal of PCA is to reduce the dimensionality of the original data by projecting them into a lower dimensionality space without a significant loss of information. This can be done by identifying the directions that explain the maximum variation of the data. The PCA method captures the variability of a process by monitoring the T^2 metric on the new PCA components or by monitoring the residuals (Q chart) of the PCA model [4]. In case of non-linear processes, kernel principal component analysis (KPCA) was used to handle non-linearity with the help of kernel functions [6].

Another multivariate method based on statistical learning approaches is the one-class Support Vector Machines (1-SVM) [7], which is a variant of the original Support Vector Machine (SVM) algorithm [8]. 1-SVM is a useful and popular tool used for anomaly detection. A static model based on 1-SVM method and the SVM-recursive feature elimination algorithm (RFE-SVM) [9] was used in [10] for fault detection in a semiconductor etch process, and in chemical engineering simulation problem. It has been shown that 1-SVM method is an efficient method for fault detection in both domains. Moreover, the 1-SVM technique performed better than PCA. Even in non-linear cases, simulation experiments showed that 1-SVM technique outperformed the KPCA method.

However the SVM-RFE algorithm requires a huge computational time since the number of SVM models to be trained is $O(p^2)$, where p is the dimension of variable space. In our study, the variable space is characterized by several electrical parameters (hundreds or thousands). High dimensional variable space restricts the use of the SVM-RFE algorithm. Moreover, as part of the training stage at each iteration of a real-time application, this algorithm would not be computationally useful, especially when we use a short temporal moving window to update the detection algorithm.

To overcome this problem, we have developed in [11] a new filter technique selecting the most relevant features (electrical parameters). This technique is based on the Median Absolute Deviation method denoted by MADe [12], a robust approach for detecting univariate outliers. The key idea is to use the MADe method to determine the percentage of outlier in each parameter. Then parameters with a percentage of outliers exceeding a predefined threshold will be potential discriminative features. We denote this method by MADe-FS (MADe for Feature Selection).

The remainder of the paper is structured as follows. First, our main contributions in this work are mentioned in the Section II. In Section III we recall the one-class support vector machine method. Then, our filter method MADe-FS which selects the most informative parameters is also

recalled in Section IV. Section V describes our real-time detection system according to two proposed scenarios for updating moving window. A short description of Hotelling's T^2 test which is the basis of an alternative detection system is given in Section VI. Before concluding, Section VII serves as an application of our system on a two real-world semiconductor datasets.

II. MAIN CONTRIBUTIONS

In our work [11], we have considered the problem of detecting abnormal wafers in semiconductor using electrical measurements. We have developed a static model for fault detection based on 1-SVM method for anomaly detection and our filter method MADe-FS for selecting the most relevant electrical parameters.

In this work, we consider the problem of real-time fault detection, becoming increasingly important in semiconductor domain. We develop a dynamic model which shares the same approaches of classification and feature selection as in our static model. Our dynamic model consists of updating the MADe-FS method and the 1-SVM algorithm at each update of the moving temporal window. We propose two scenarios for updating this window, and we explain our technique used to optimize the initial choice of model parameters and their updating strategy.

As an alternative system of real-time detection, we implement a similar dynamic model based on PCA method to reduce dimension and model the normal behavior, and Hotelling's T^2 statistic as multivariate control chart. Parameters of this model is selected and updated under the same strategy used in our developed dynamic model.

At our knowledge, this work is the first one to implement a real-time fault detection system in semiconductor domain, and at the same time the first one to develop a dynamic model based on the 1-SVM method. This model is applied on high dimensional data consisting of hundreds of variables while previous works on fault detection in industrial processes considered data with tens of variables.

III. ONE-CLASS SUPPORT VECTORS MACHINES

Support Vector Machine (SVM) [13] is as an effective learning algorithm for binary classification. This algorithm aims to find an optimal hyperplane to separate the two classes of training data.

An extension of SVM, called one-class SVM (1-SVM), was subsequently proposed in [7] to handle one-class classification problem. The 1-SVM strategy is to find an optimal hyperplane in a feature space separating the training data (positive samples) from the origin (considered as negative samples) with maximum margin (the distance from the hyperplane to the origin).

Given a training dataset of n positive samples (normal wafers) $\{x_1,\ldots,x_n\}$ where each $x_i \epsilon R^p$ is described by a vector of p features (electrical parameters). Each x_i is first transformed via a feature map $\varphi: R^p \dashrightarrow F$ where F is a high (possibly infinite) dimensional Hilbert space generated by a positive-definite kernel K. The kernel function corresponds to an inner product in the feature space F through $K(x, x') =$

$\varphi(x) \cdot \varphi(x')$.

The 1-SVM algorithm finds in the feature space a hyperplane $H \{z \in F; w \cdot z = \rho\}$ that separates the cluster of normal samples from the origin. $w \epsilon F$ is the normal vector defining H. The margin is equal to $\rho/\|w\|$. The one-class SVM requires solving the following quadratic optimization problem:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + \frac{1}{vn} \sum_{i=1}^{n} \xi_i - \rho$$

$$st \quad w \cdot \varphi(xi) \geq \rho - \xi_i, \xi_i \geq 0, i = 1, \ldots, n. \quad (1)$$

ξ_i's are slack variables introduced to allow misclassification for some points, and $v \in [0, 1]$ is a free parameter controlling the impact of the slack variables, i.e. the fraction of training data which are allowed to fall wihtin the margin. In fact, it can be shown that v is an upper bound on the fraction of training errors [7].

The dual problem, to be maximized, is given by:

$$\min_\alpha \frac{1}{2} \sum_{ij} \alpha_i \, \alpha_j \, K(x_i, x_j)$$

$$st \quad 0 \leq \alpha_i \leq \frac{1}{vn}, \sum_i \alpha_i = 1. \quad (2)$$

The data x_i with non-zero α_i are the so-called support vectors. They are the training data that determine the separating hyperplane. It can also be shown that v lower bounds the fraction of support vectors [7].Once the optimal values of the parameters are found, one can classify the new data (new wafers) according to the decision function

$$g(x) = sgn(\sum_{i \, \epsilon \, sv} \alpha_i \, K(x_i, x) - \rho), \quad (3)$$

where sv is the set of the support vectors' indices.

In practice, the 1-SVM has been successfully applied with the RBF kernel $K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|2)$ where γ is a parameter that controls the width of the kernel function. After many experiments in which we have tested many values for γ ($\gamma = 1/mp$, with $m \in \{0.5, 1, 2, 3, 4, 5\}$), results have showed that best performance of 1-SVM algorithm is obtained for $m > 1$, and this performance is not very sensitive to the kernel parameter. Hence fine-tuning of the parameter is not required. We set $\gamma = 1/5p$.

IV. OUR FILTER METHOD

In machine learning and statistics, feature selection is the process of selecting an optimal subset of relevant features in order to obtain good classification performances.

To achieve the task of feature selection, we use our filter approach based on MADe method, which is a robust univariate outlier detection method. Before presenting this method, we first introduce the Maximum Absolute Deviation (MAD) [14] of a variable $x_j \epsilon R^n$ ($j = 1, \ldots, p$):

$$\text{MAD}(j) = \text{median}_{i \, \epsilon \, [1,n]} (|x_{ij} - \text{median}(x^j)|). \quad (4)$$

MAD is a robust estimator of the spread in a data, similar to the standard deviation. When the MAD value is scaled by a factor of 1.483, it represents a consistent estimator of the standard deviation in a normal distribution [12]. This scaled

MAD value is the MAD_e:

$$\mathrm{MAD}_e = 1.483 \times \mathrm{MAD}. \qquad (5)$$

The MAD_e method is defined as follows:

$$LL_j = \mathrm{Median}(j) - 3 \times \mathrm{MAD}_e(j)$$
$$UL_j = \mathrm{Median}(j) + 3 \times \mathrm{MAD}_e(j), \qquad (6)$$

where LL_j and UL_j are respectively the lower and upper limits for the variable j.

The MADe approach is similar to the Standard Deviation (SD) method that considers the observations outside the interval $[\bar{x}\pm3\sigma]$ as outliers, where x̄ and σ are respectively the empirical mean and standard deviation for a univariate samples. However, the median and MAD_e are employed instead of the mean and the standard deviation. Since this approach uses two robust estimators, it is largely unaffected by the presence of extreme values in the data set.

Thus the percentage of outliers OOL_j (Out Of Limit) of the variable x_j represents the proportion of data outside the interval determined by the lower and upper limits of the MADe method. Therefore we have:

$$OOL_j = \frac{\# i \,\epsilon\, [LL_j, UL_j]}{n} \qquad (7)$$

Finally, the Subset of Relevant Variables (SRV) contains variables for which the percentage of outliers exceeds the threshold θ_q (cf. Eq. 8). θ_q is defined as the quantile of order q of the values in the vector OOL=(OOL$_1$,…, OOLp).

$$SRV = \{x^j, \mathrm{OOL}_j > \theta_q\} \qquad (8)$$

To conclude, our filter method consider the top $100(1-q)$ % outlying variables as the most relevant electrical parameters for the classification task.

V. REAL TIME DETECTION SYSTEM

The motivation behind the development of a real-time detection system is to use the MADe-FS and 1-SVM approaches for in-line testing in the context of industrial application. This system aims to detect in real-time abnormal wafers using a recent history of electrical measurements. In the following, we denote our model of feature selection and classification by 1-SVM.FS (one-class SVM with Feature selection). This model consists of determining firstly the most relevant features in the training data using our filter method MADe-FS, and secondly applying the 1-SVM algorithm on the subset of relevant features.

Our detection system is based on three major steps:

1) Selection of a correct performance reference data set, representing the normal operating behavior
2) Real-time data updating through a moving window, to obtain a real-time procedure.
3) 1-SVM.FS application to the updated real-time data.

So we first define the reference correct performance dataset, representing a well-behaved operating condition. For this, we select from the historical database of considered products, a set of operational positive samples (normal wafers) corresponding to a nominal condition of processes. Concerning the reference data size, a large data set increases the detection reliability. Hence reference data must be large enough allowing us to define a normal region which encompasses a wide variety of positive samples.

The correct performance dataset is used as a training set to build a model describing the normal behavior of the process. When a new lot (group of 25 wafers that run together all processing steps) arrives, the 1-SVM.FS model trained on the correct performance dataset is used to test whether each of 25 wafers is normal or abnormal. The tested lot will join the initial training set while oldest lot in this set will be removed or maintained depending on the used scenarios explained below. Thus a new training set is formed. 1-SVM.FS model is retrained on the updated training set and will be used to predict the operating state of the next 25 new wafers. 1-SVM.FS model is retrained on the updated training set and will be used to predict the operating state of the next 25 new wafers. 1-SVM.FS model is retrained on the updated training set and will be used to predict the operating state of the next 25 new wafers. This procedure is repeated with the arrival of each new lot. A general view of our detection system is presented in the Fig. 1.
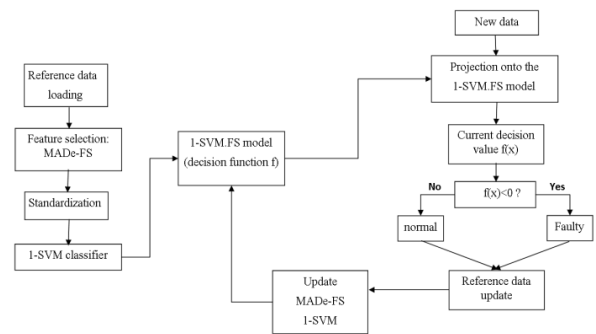

Fig. 1. Schema of our real-time detection system based on 1-SVM.FS dynamic model.

As we described before, the basic 1-SVM.FS model was made dynamic by updating the database through a moving window. We consider 2 scenarios reflecting two updating modes of the moving window:

- Scenario 1 (increased length): with this scenario, the tested lot at each iteration is added to the existing training set without removing old data. So 1-SVM.FS model is updated according to a moving window of increased length. Since normal behavior keeps evolving, we have decided to remove at once some old data from the increased training set after a defined period Δt. Δt depends on the volume production of the considered product(s).
- Scenario 2 (fixed length): during the real-time operation, the window still maintains the length of the correct performance dataset and operates as a First-In-First-Out (FIFO) shift-register, discarding old data and including new ones.

The two scenarios are illustrated respectively in the Fig. 2 and the Fig. 3.

Recall that 1-SVM.FS model requires setting the parameter ν (the threshold in 1-SVM algorithm) and two hyperparameters: the order q of the threshold θ_q in feature selection method, and the kernel parameter γ. Consequently some kind of model selection (parameter search) must be done.

To accomplish the model selection task, a validation set containing normal data contaminated by some abnormal wafers is needed. It is used to identify good (q, v) so that the classifier can accurately predict unknown data (i.e. testing data). A "grid search" on q and v is performed. 1-SVM.FS model is built on training set using various pairs of (q, v) values. For each pair, samples from validation set are projected onto the trained 1-SVM.FS model. Then Detection Rate and the False Alarms Rate (cf. Section VII-A) are computed. The pair that optimizes these two performance measures is picked. More precisely, the best pair (q, v) is the one giving the optimal compromise between maximizing the Detection Rate and minimizing the False Alarms Rate. The selected pair is used at each update of the 1-SVM.FS model.

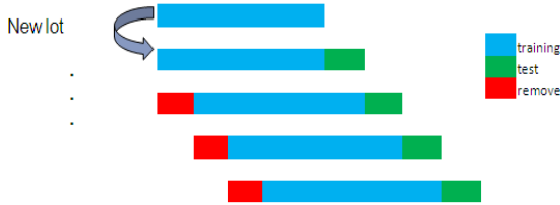Fig. 2. Real-time moving window using scenario 1.

Fig. 3. Real-time moving window using scenario 2.

VI. Hotelling's T^2 Test

To make our study comparable to previous studies, we have investigated the Hotelling's T^2 test. Hotelling's T^2 statistic provides an indication of novel variability within the model space.

The principle of this test is to use PCA method to model the behavior of the normal samples. Anomalies are then detected by comparing the behavior observed with that given by the PCA model. Having established a PCA model of the positive training data, testing data are projected onto this model, and Hotelling's T^2 statistic can be computed based on the first k principal components of the model. The T^2 statistic for a sample x_i is:

$$T_i^2 = t_i^T \wedge^{-1} t_i = x_i^T p_k \wedge^{-1} p_k^T x_i \qquad (9)$$

where $t_i = P_k^T x_i$ is the orthogonal projection of the data x_i into the model subspace defined by the k first principal components, and Λ is a diagonal matrix containing the first k eigenvalues of the covariance matrix of the positive training data.

A threshold T^2_α can be obtained using the Fisher distribution. If $T^2_i > T^2_\alpha$, the sample is categorized as abnormal, and normal otherwise. For further details on fault

detection based on PCA readers are advised to read the literature [4].

To choose k, we use the Cumulative Proportion of Variance (PCV):

$$\text{PCV}(k) = 100 \times \frac{\Sigma_{j=1}^{k} \lambda_j}{\Sigma_{j=1}^{p} \lambda_j},$$

where $\lambda 1, \ldots, \lambda p$ are the eigenvalues sorted in descending order. Thus we retain the first k components that account for a predefined percentage of the variance in the data:

$$k = \arg\min_u \{\text{PCV}(u) \geq \beta \}.$$

For example, if we set $\beta = 0.8$ we retain the minimal number of components that preserve 80% of the information in the data.

Detection system based Hotelling's T^2 test is dynamically the same as our system. The data and model update is performed at the level of 25 wafers (each new lot) following the proposed two scenarios.

VII. Application

Our experimental goal was to assess the ability of our detection system to detect abnormal wafers. It is also important to minimize false alarms rate as they cause unwarranted interruption in plant operation. Let us first introduce the performance measures used in our study.

TABLE I: Confusion Matrix of Metrics Used in Performance Measures

True class vs Decision	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

A. Performance Measures and Data

In order to evaluate and compare the results obtained from the different methods, we used two performance criteria: Detection Rate (DR) and False Alarms Rate (FAR). Detection Rate quantifies the percentage of data predicted to be negative by the classifier that are actually negative; False Alarms Rate quantifies the percentage of data predicted to be negative by the classifier that are actually positive. These two measures are computed using the four metrics described in the Table I as follows:

$$DR = \frac{TN}{TN + FN}$$

$$FAR = \frac{FP}{FP + TP}$$

We notice that the resulting false alarms rate in the context of application of real-time detection system over a production period represents the average of false alarms rates obtained when testing separately each of all lots that have to be tested. Furthermore, the FAR-DR curve is suitable for evaluating classifiers by integrating their performance over a range of decision thresholds. It depicts the relation between DR (x-axis) and FAR (y-axis) varying the range of thresholds. The lower the misclassification error of a classifier, the closer the corresponding point is to the upper right-hand corner of the ROC curve.

The real-time detection system proposed in this paper has been tested on two real-world industrial datasets. Each dataset consists of wafers corresponding to one or more products of a certain technology over months of production. Each wafer is described by a certain number of electrical parameters. We give the percentage of abnormal wafers in each dataset. The description of these two datasets is given in Table II. 1-SVM.FS and Hotelling's T^2 detection systems are investigated under the two scenarios in both datasets, in order to prove again the efficacy and superiority of our detection system. Ideally, we want high DR (to detect most of the abnormal wafers) and a low false alarms rate (to avoid mistakenly classifying normal wafers as abnormal).

TABLE II: Description of the Real World Industrial Data Used in Our Study

Data	Production time	Nb of parameters	% of abnormal wafers
dataset 1	2 months	756	1.75
dataset 2	4 months	1062	0.5

1) Dataset 1

In this experiment, the correct performance data is formed using 300 normal wafers. The validation set consists of 100 wafers of which 6 wafers are abnormal. We have trained our 1-SVM.FS model on the correct performance data using various pairs of values for the feature selection hyperparameter q and the threshold v. We consider respectively 6 and 20 values for q and v, as follows:

$$q \epsilon \{0.25, 0.4, 0.5, 0.6, 0.75, 0.9\},$$

$$v \epsilon \{0.01, 0.02,\ldots, 0.19, 0.2\}.$$

Samples from the validation set are then predicted using each of 120 learned models. The Detection Rate and the False Alarms Rate are computed for each prediction. We have selected the pair that optimizes simultaneously these two performance measures. Here we have retained $q=0.75$ and $v=0.16$ and we have obtained a DR equal to 100% and FAR equal to 14.21%.

Similarly, we have selected for the Hotelling's T^2 test the best pair (β, α) (cf. Section VI) by taking $\beta \epsilon \{0.75, 0.8, 0.85, 0.9\}$ and considering the same range of values of v for α. The optimal performance is obtained for $\beta=0.75$ and $\alpha=0.2$, where DR and FAR are respectively equal to 66.67% (4 among 6 abnormal wafers) and 17.36%.

After defining the correct performance data set and selecting the optimal parameters for 1-SVM.FS and Hotelling's T^2 models, we now proceed to the real-time detection by applying both of models to the real-time data. The real-time data are updated at each arrival of a new lot through a moving window in order to obtain a real-time procedure. The two models are also updated. The updates through the moving window follow one of two defined scenarios: scenario 1 (increased length) and scenario 2 (fixed length).

Next, we focus on comparing the performance of the two real-time detection systems based on 1-SVM.FS and Hotelling's T^2 dynamic models using the two scenarios. The results are reported in Table III. For both scenarios, the Hotelling's T^2 system has shown poor performance in detecting abnormal wafers (DR=65.22%), while 1-SVM.FS system has been able to detect 95.65% of abnormal wafers. In

addition, we have obtained lower false alarms rate using our detection system. For both detection systems, scenario 1 reduces false alarms compared to scenario 2.

TABLE III: Performance of 1-SVM.FS and Hotelling's T^2 Systems on the Dataset 1

Moving window	Detection system	Detection Rate	False Alarms Rate
Scenario 1	1-SVM.FS	95.65	12.89
	Hotelling's T^2	65.22	13.43
Scenario 2	1-SVM.FS	95.65	19.25
	Hotelling's T^2	65.22	19.85

To confirm this hypothesis, FAR-DR curve is plotted in the Fig. 4 to study the behavior of our detection system regarding the two different scenarios, over the same range of v defined above. It is clear that scenario 1 gives a significant reduction interm of false alarms compared to scenario 2. This is due to the increased size of its moving window where a new lot is added to the training database at each update. In fact one-class SVM requires many more positive training data to give an accurate decision boundary because its support vectors come only from the positive data. However scenario 2 tends to detect more quickly abnormal wafers (i.e. for any value of v, scenario 2 has higher or the same DR than scenario 1). The short fixed window in scenario 2 has a more efficient updating strategy and contains fewer abnormal wafers in the moving training dataset, which improves the performance of 1-SVM algorithm since the latter requires normal wafers to learn the classifier.
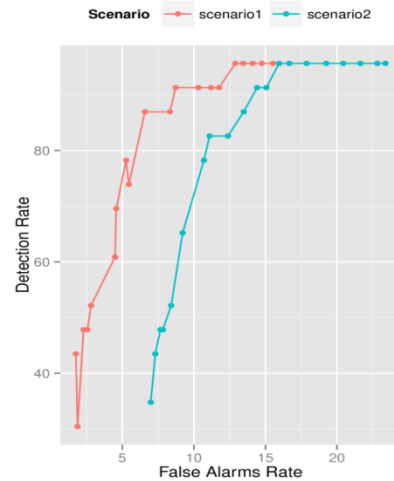

Fig. 4. FAR-DR curve comparing performances of 1-SVM.FS detection system using the two proposed scenarios.

Note that, in the first experiment considering only two months of production, we did not remove old data in the actual training set after the Δt period for the scenario 1, as has been recommended in Section V. This action takes place in the second experiment considering four months of production where we have a larger number of wafers.

A final comparison is realized between 1-SVM.FS and 1-SVM detection systems. The difference between these two systems is that the latter ignores the feature selection step used by the former. Another FAR-DR curve is plotted in the Fig. 5 illustrating this comparison. From this curve, a very important improvements achieved by applying our feature selection method MADe-FS. These improvements were

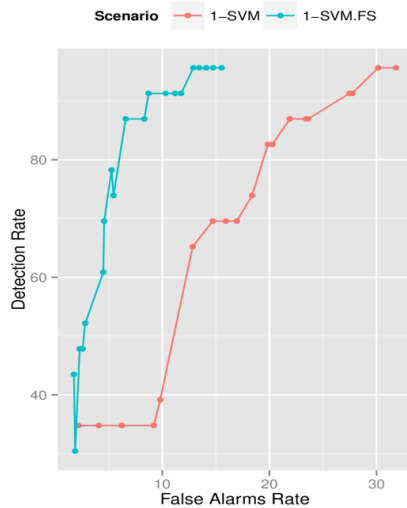observed on each of the two performance measures (DR and FAR).

Fig. 5. FAR-DR curve showing the importance of our filter method MADe-FS to improve the performance of the 1-SVM classifier, according to scenario 1.

2) Dataset 2

Dataset 2 contains wafers from another category of products collected over four months of production. This dataset has higher dimensional space and lower percentage of abnormal wafers, compared to the first dataset. We set to 500 the size of the correct performance data. The validation set contains 2 abnormal wafers among of 100.

Following the same procedure used in dataset 1 for selecting optimal parameters, we have retained $(q, v)=(0.75, 0.04)$ for 1-SVM.FS model and $(\beta, \alpha)=(0.8, 0.01)$ for Hotelling's T^2 model. We set Δt to 2 months.

Table IV summarizes the performances achieved by the two systems under the two different scenarios. The results reveal a degree of similarity between the performances of both systems. High performance was obtained using both systems. Observations resulting from the comparison of two scenarios in dataset 1 are confirmed in dataset 2. Scenario 1 has lower false alarms rate, while scenario 2 detect more effectively abnormal wafers.

TABLE IV: Performance of 1-SVM.FS and Hotelling's T^2 Systems on the Dataset 2

Moving window	Detection system	Detection Rate	False Alarms Rate
Scenario 1	1-SVM.FS	83.33	5.89
	Hotelling's T^2	83.33	6.34
Scenario 2	1-SVM.FS	91.67	8.62
	Hotelling's T^2	91.67	9.12

VIII. Conclusion

In this paper, we proposed a new real-time fault detection system based on the machine learning 1-SVM algorithm and our filter method for feature selection. A dynamic detection was realized by updating the database following two proposed scenarios. The efficacy of our system has been demonstrated using two real-world industrial datasets. For both scenarios, results from the two datasets showed that our system could detect most of the abnormal wafers with an admissible percentage of false alarms. In addition, our system outperformed the detection system based on the Hotelling's T^2 test in the dataset 1, and similar performance was obtained in dataset 2 with slightly lower rate of false alarms.

References

[1] T. Kourti, J. Lee, and J. Macgregor, "Experiences with industrial applications of projection methods for multivariate statistical process control," *Computers and Chemical Engineering*, vol. 20, pp. S745–S750, 1996.

[2] A. Cinar and C. Undey, "Statistical process and controller performance monitoring: a tutorial on current methods and future directions," in *Proc. American Control Conference*, 1999, vol. 4, pp. 2625–2630.

[3] B. Wise, N. Gallagher, S. Butler *et al.*, "A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process," *Journal of Chemometrics*, vol. 13, pp. 389–422, 1999.

[4] D. Garcia-Alvarez, M. Fuente, P. Vega, and G. Sainz, "Fault detection and diagnosis using multivariate statistical techniques in a wastewater treatment plant," in *Proc. 7th IFAC International Symposium on Advanced Control of Chemical Processes*, 2009, pp. 952–957.

[5] F. Baggiani and S. Marsili-Libelli, "Real-time fault detection and isolation in biological wastewater treatment plants," *Water science and technology*, vol. 60, no. 11, pp. 2949–2961, 2009.

[6] J. Lee, C. Yoo, S. Choi, P. Vanrolleghem, and I. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chemical engineering Science*, vol. 59, no. 1, pp. 223–234, 2004.

[7] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.

[8] V. Vapnick, "Sv machines for pattern recognition," *Statistical Learning Theory*, John Wiley Sons, pp. 496–498, 1998.

[9] A. Rakotomamonjy, "Variable selection using svm based criteria," *J. Mach. Learn. Res.*, vol. 3, pp. 1357–1370, Mar. 2003.

[10] S. Mahadevan and S. Shah, "Fault detection and diagnosis in process data using one-class support vector machines," *Journal of Process Control*, vol. 19, pp. 1627–1639, 2009.

[11] A. H. Hassan, S. Lambert-Lacroix, and F. Pasqualini, "A new approach of one class support vector machines for detecting abnormal wafers in semi-conductor," in *Proc. Fourth Meeting on Statistics and Data Mining, ser. MSDM '13,* 2013, pp. 35–41.

[12] S. Burke, "Missing values, outliers, robust statistics and non-parametric methods," *Statistics and Data Analysis*, vol. 2002, 2001.

[13] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, Jun. 1998.

[14] F. R. Hampel, "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, 1974.

Ali Hajj Hassan is a Ph.D. student at Grenoble University, France. He works as an engineer in the Process Control Department at STMicroelectronics Crolles. He received the applied mathematics and statistics degree from University of Joseph Fourier, Grenoble, France, in 2010. His Ph.D. subject is about multivariate statistical approaches for detecting abnormal wafers in semiconductor.