Identifying Systematic Spatial Failure Patterns through Wafer Clustering

Mohamed Baker Alawieh, Fa Wang and Xin Li
Electrical and Computer Engineering Department
Carnegie Mellon University
Pittsburgh, USA
{malawieh, fwang1, xinli}@ece.cmu.edu

***Abstract*— In this paper, we propose a novel methodology for detecting systematic spatial failure patterns at wafer level for yield learning. Our proposed methodology takes the testing results (i.e., pass or fail) of a number of dies over different wafers, cluster all these wafers according to their failures, and eventually identify the underlying spatial failure patterns. Several novel machine learning algorithms, including singular value decomposition, hierarchical clustering, dictionary learning, etc., are developed in order to make the proposed methodology robust to random failures. The efficacy of our proposed approach is demonstrated by an industrial data set.**

Keywords—wafer clustering; defect patterns; singular value decomposition; random failures.

I. INTRODUCTION

The challenges associated with designing and manufacturing of leading edge integrated circuits (IC) have increased with the complexity of chip functionalities. Such complex functionalities have been made possible by the continuous drive towards scaling IC technologies [1]. However, with such scaling, catastrophic defects and process variations stand out among the most important factors limiting the product yield of IC designs [2]-[5].

In order to improve the product yield in the IC design cycle, it is important to identify the underlying factors that contribute most to the yield loss [6]-[7]. As reported in the literature, different wafers may have remarkably different spatial signatures for a given measurement [8]. This in fact is due to the presence of different underlying defect sources contributing to the different spatial signatures. The presence of these variation sources is demonstrated by studying the contribution of different user-defined variation patterns on the die-level performance [9]-[10]. Therefore, to help yield learning, we need to identify the systematic spatial failure patterns among the wafers [6]. This goal is accomplished by partitioning the wafers into groups with similar spatial signatures. Such partitioning will help process engineers focus on the failure causes associated with the significant yield loss [7].

Automatically grouping wafers of similar spatial signatures can be formulated as a clustering problem. Although clustering analysis has been extensively studied in the statistics community, the task of wafer clustering targeted here has its unique characteristics. Among the important characteristics of this problem is the binary nature of the measurements (i.e., pass or fail) for most digital test items. Similar problem has been tackled in [11], where the clustering is done based on measurements of continuous performance metrics. In practice, many measurements are done to verify whether chips on a wafer provide the correct functionality or meet a required design specification; hence, the result of such testing is binary. For this reason, we will consider in our approach that the available test data is in binary format.

Another important challenge associated with this problem is the presence of random defects and variations that can mask the underlying systematic failure pattern; thus, making the detecting task non-trivial [12]-[14]. Moreover, few wafers can have abnormal signatures that may be due to equipment malfunction. It is important to detect these special wafers as outliers and avoid including them in the desired clusters [11].

Given these problem features, we aim to identify the systematic wafer-level spatial pattern while significant random defects and variations exist. To achieve this goal, we propose a new methodology for clustering of wafer spatial signatures. The proposed method consists of three major steps. First, singular value decomposition (SVD) is performed on the available binary testing data, and then data is retrieved by using a chosen number of the important singular vectors. This step serves as a feature extraction method to reduce the impact of random defects and variations prior to performing the required clustering. Furthermore, this step will help in detecting outliers among wafers. Next, hierarchical clustering is applied on the retrieved data to generate a set of possible clustering results where the optimal number of clusters is chosen to get the final clustering [15]. Finally, to further reduce the effect of random defects and variations, a dictionary learning algorithm is adopted from binary data compression [17] and used to reconstruct the binary signatures of the resulting clusters. The reconstruction process includes only the important dictionary entries that are expected to capture the systematic spatial failure patterns.

The remainder of this paper is organized as follows. In Section II, we present the details of the proposed wafer clustering algorithm. The efficacy of the proposed method is demonstrated using an industrial example in Section III. Finally we conclude in Section IV.

II. PROPOSED METHOD

Our proposed algorithm for identifying systematic spatial failure patterns is composed of three major steps: (i) SVD, (ii) hierarchical clustering, and (iii) binary dictionary learning.

A. *Singular Value Decomposition*

Our goal is to capture the systematic spatial failure patterns

978-1-4799-5341-7/16/$31.00 ©2016 IEEE

in the presence of significant random defects and variations. Singular Value decomposition (SVD) comes to play here as a means to extract important features of the systematic spatial signatures, while removing the random variations.

SVD is applied on the d-by-N matrix $\mathbf{X}$ containing the binary test data obtained from the wafers [15]:

$$\mathbf{X} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T , \quad (1)$$

where d is the number of dies per wafer, N is the total number of wafer, $\mathbf{U}$ is a d-by-d matrix whose columns represent the left singular vectors, $\mathbf{V}$ is an N-by-N matrix whose columns represent the right singular vectors, and $\mathbf{\Sigma}$ is a d-by-N diagonal matrix whose diagonal elements represent the singular values.

The singular values in $\mathbf{\Sigma}$ reflect the importance of the corresponding singular vectors in $\mathbf{U}$. Therefore, we can reduce the random effects in the test data by considering the first k important singular vectors that carry the systematic pattern signatures. Namely, we want to choose k important singular vectors to represent the data.

To get the optimal value of k, its value is swept in an increasing order and at each step the total reconstruction error is computed as:

$$E_k = \left\| \mathbf{X} - \mathbf{U}_k \cdot \mathbf{\Sigma}_k \cdot \mathbf{V}^T \right\|_F , \quad (2)$$

where E_k is the value of the total error when k singular values are considered, $\mathbf{U}_k$ is the matrix containing the first k columns of $\mathbf{U}$, $\mathbf{\Sigma}_k$ is the matrix containing the first k rows of $\mathbf{\Sigma}$, and $\|\bullet\|$ represents the Frobenius norm of a matrix.

The optimal value of k is chosen as the value where adding more singular vectors will not significantly improve the total reconstruction error. This implies that any additional singular vector will only affect few dies on some wafers, thus it does not represent a systematic pattern but rather tries to model the random effects.

Once the optimal value of k is chosen, the data set is reconstructed as:

$$\mathbf{X}_K = \mathbf{U}_K \cdot \mathbf{\Sigma}_K \cdot \mathbf{V}^T , \quad (3)$$

where K is the optimal value of k, , $\mathbf{U}_K$ is the matrix containing the first K columns of $\mathbf{U}$, and $\mathbf{\Sigma}_K$ is the matrix containing the first K rows of $\mathbf{\Sigma}$.

B. Hierarchical Clustering

We propose to group all wafers into a small number of clusters where the wafers in the same cluster share the same spatial failure pattern. As such, the failure patterns can be robustly extracted from these clusters. Since the data set is expected to contain outlier wafers with abnormal spatial signatures, hierarchical clustering is used so that the clustering results are not strongly biased by the outliers [11], [15]. The hierarchical cluttering algorithm tries to build clusters in a greedy manner starting from individual clusters, and iteratively, merging the two clusters with the shortest distance.

However, the definition of the distance between two clusters is not unique. In our implementation, we use the complete-link version of hierarchical clustering which defines the distance between two clusters as the maximum distance between two wafers belonging to these clusters. Physically, this means that all wafers in the same cluster are within a small distance from each other [15].

Another important implementation issue is to choose the number of clusters. In practice, hierarchical clustering algorithm will provide at each merging step the current distance between the two merged clusters. Based on this measure, we use the modified L-Method described in [11],[15] to choose the optimal number of clusters.

The modified L-method generates a curve showing the merging distance versus the number of clusters at each merging step. While this curve presents a decreasing trend with the increase in the number of clusters, it usually has a sharp transition at the optimal clustering setup. The graph will show two different trends of decaying before and after the transition point. At this point, all merged clusters are close to each other and the distance between any two of the current clusters is relatively high. Hence, capturing this transition point will reveal the optimal number of clusters [11], [15]-[16].

Since the wafer data is expected to contain abnormal wafers, single-wafer clusters may appear after clustering. These wafers are totally different from the formed clusters; hence hierarchical clustering will not merge them due to the high merging distance. The single-wafer clusters actually represent outlier wafers that are different from the systematic patterns present across the wafers; therefore, they are not used in the next steps since they do not carry any systematic failure pattern.

After the final clusters are formed, we are interested in obtaining one binary signature per cluster that represents its systematic spatial pattern. To achieve this goal, the spatial wafer patterns in the same cluster are averaged to build a single binary signature for each cluster.

C. Dictionary Learning

The binary group signatures carry the systematic spatial patterns of the wafers. However, random variations and defects may be still present in these patterns. Therefore, a dictionary learning step serves as a second stage of filtering out these random effects.

We assume that the number of clusters obtained from hierarchical clustering is c, and $\mathbf{y}_i$ is the binary spatial signature corresponding to the ith cluster ($i = 1, 2, \cdots, c$). Our goal is to form a dictionary that can represent these data using a compact set of basis vectors and coefficients [17]. Next, in an approach similar to choosing the important singular values in Section II-A, only important dictionary entries are used to reconstruct the spatial signatures.

In the first step, the available group signatures are considered as the candidates from which dictionary entries are chosen. Given the c available signatures, we want to choose the first entry in the dictionary as the signature that can achieve the lowest reconstruction error. Then, we iteratively choose the signatures that will reduce the reconstruction error of the residual.

Since we are trying to build a binary dictionary, the associated coefficients are also binary. These coefficients are computed one at a time. In other words, when one dictionary entry is chosen, the coefficients associated with this entry are computed before choosing the next entry.

Initially, the dictionary is empty, and we initialize the residual vectors $\mathbf{r}_i$ to be equal to the vectors $\mathbf{y}_i$ ($i = 1, 2, \cdots, c$). Hence, we compute the reconstruction error associated with each candidate as [17]:

$$E_i = \sum_{j=1}^{c} \mathrm{wt}\left[\mathbf{r}_j \oplus \left(\alpha_{i,j} \otimes \mathbf{r}_i\right)\right] \quad \left(i = 1, 2, \cdots, c\right), \tag{4}$$

where $\oplus$ is the XOR function, $\otimes$ is the binary multiplication, the weight function wt(•) returns the number of ones in a binary vector, and $\alpha_{i,j}$ is a coefficient defined as [17]:

$$\alpha_{i,j} = \begin{cases} 1, & \mathrm{wt}\left(\mathbf{r}_j \oplus \mathbf{r}_i\right) < \mathrm{wt}\left(\mathbf{r}_j\right) \\ 0, & \mathrm{wt}\left(\mathbf{r}_j \oplus \mathbf{r}_i\right) \geq \mathrm{wt}\left(\mathbf{r}_j\right) \end{cases}. \tag{5}$$

Based on (4)-(5), the candidate with the lowest reconstruction error is chosen as the first dictionary entry and the coefficients associated with it are saved [17].

At this stage, we have chosen the signature that represents the first dictionary entry; however this signature may still suffer from random variations and defects. As another filtering step, the residual vectors with non-zero coefficients associated with this entry are averaged and the bits of the resulting mean are rounded (to ‘0’ or ‘1’). The resulting binary vector is the actual entry saved in the dictionary.

Next, the residual is updated to:

$$\mathbf{r}_i = \mathbf{r}_i \oplus \left(\alpha_i^* \otimes \mathbf{r}^*\right) \quad \left(i = 1, 2, \cdots, c\right), \tag{6}$$

where $\mathbf{r}^*$ is the new dictionary entry and α^*_i ($i = 1, 2, \cdots, c$) is the coefficient associated with it.

In the following iterations, the reconstruction error is computed again according to (4)-(5), a new entry is added to the dictionary, and the residuals are updated according to (6). When all residuals are zero, the dictionary is complete and the associated coefficients are also available [17].

The goal behind this dictionary learning step is to filter out the random variations and defects and capture the systematic spatial patterns present in the data. Hence, once again, when reconstructing the group signatures we will only choose the first few important dictionary entries to represent the data. These vectors are expected to carry the systematic patterns of interest.

However, the way the dictionary is built may result in highly correlated dictionary entries especially when the systematic patterns are overlapping. For this reason, before reconstructing the signatures, a final dictionary update is done to remove the correlation between the entries.

After choosing the first l important entries in the dictionary, the correlation between each entry and all other ones is removed, starting from the last entry in the dictionary. This is done through the update equation:

$$d_i = \begin{cases} d_i, & \mathrm{wt}\left(\left(d_j \oplus d_i\right) \oplus d_j\right) \leq \mathrm{wt}\left(d_j \oplus d_i\right) \\ d_i \oplus d_j, & \mathrm{wt}\left(\left(d_j \oplus d_i\right) \oplus d_j\right) > \mathrm{wt}\left(d_j \oplus d_i\right) \end{cases} \quad \left(j = 1, 2, \cdots, l\right), \tag{7}$$

where $\mathbf{d}_i$ is the ith entry in the dictionary, and this update is done for all entries as the index i goes from l to 1.

Once all the dictionary entries are updated, the coefficients are recomputed according to (5)-(6) for each of the entries of the dictionary starting from the coefficients corresponding to the first entry. Algorithm 1 summarizes the major steps for the dictionary learning stage. Finally, the group signatures are reconstructed based on the formed dictionary and coefficients. These signatures carry the systematic spatial patterns of the wafer data.

Algorithm 1: Dictionary Learning

```
1. Initialize $\mathbf{r}_i$ to $\mathbf{y}_i$ ($i = 1, 2, \cdots, c$).
2. While $\exists\ \mathbf{r}_i \neq 0$ ($i = 1, 2, \cdots, c$)
3.     Choose one signature according to (4)-(5).
4.     Update the candidate by averaging the signatures with non-zero coefficients.
5.     Update the residuals according to (6).
6. End While
7. Choose $l$ important dictionary entries.
8. Remove correlation between dictionary entries using (7).
9. Update the coefficients.
10. Reconstruct the group signatures based on dictionary and coefficients.
```

D. Summary

Algorithm 2 summarizes The overall flow for our proposed method. Its efficacy will be further demonstrated by an industrial data set in the next section.

Algorithm 2: Systematic Spatial Patterns Detection

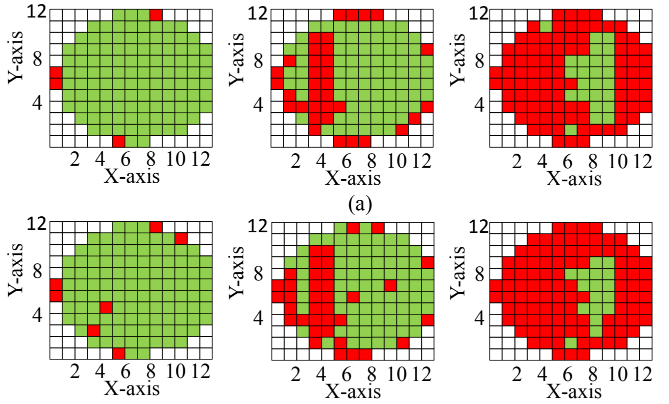
```
1. Apply SVD on $\mathbf{X}$ as described in (1) after subtracting the mean of each column.
2. Choose $k$, the number of singular vectors, reconstruct the data as shown in (3), and then add back the mean of each column.
3. Apply hierarchical clustering on the reconstructed data.
4. Obtain the optimal number of cluster using the modified L-Method.
5. Obtain binary signatures for the clusters.
6. Build the binary dictionary.
7. Reconstruct the signatures of all clusters to obtain systematic spatial patterns.
```

III. EXPERIMENTAL RESULTS

In this section, the efficacy of our proposed method is demonstrated by using the industrial measurement data collected at an advanced technology node. Each die is tested and reported as pass or fail. The data set consists of 417 wafers having 117 dies per wafer.

The proposed method for systematic failure pattern extraction is applied to the aforementioned wafer data. It results in six different clusters. Among these clusters, only three are carrying systematic spatial patterns while the other clusters are single-wafer clusters containing outlier wafers only.

Considering the three non-single-wafer clusters, each of them defines a systematic spatial failure pattern, as shown in Figure 1 (a). In addition, Figure 1 (b) shows one representative wafer map for each of these three patterns. Table 1 further shows the statistics for the three patterns. The first pattern carries a number of failed dies across all wafers. The yield of these wafers associated with the first pattern is around 97%. The two other patterns carry many failed dies, and only a small number of wafers are associated with these two patterns. The corresponding yield values are around 70% and 21% for these two patterns, respectively. Such information would be extremely helpful for the process engineers in order to improve the product yield by addressing these spatial failure patters based on their importance.



(a)

(b)

Figure 1. (a) Three systematic spatial failure patterns are identified, and (b) one representative wafer map is shown for each pattern where red and green colors represent failing and passing dies respectively.

Table 1. Statistics for the systematic spatial failure patterns across wafers

	Pattern 1	Pattern 2	Pattern 3
% of wafers	89.9%	7.9%	1.4%
Average % of failing dies per wafer	3%	30%	79%

Figure 2 shows the spatial patterns corresponding to the single-wafer clusters that are discarded and considered as outliers. It is clear that these wafers are in fact outliers as their spatial patterns are substantially different from those of the normal wafers.

The aforementioned case study based on industrial measurement data demonstrates that the proposed method is able to identify systematic spatial failure patterns in the presence of random variations and defects, and meanwhile detect the outlier wafers based on their abnormal spatial patterns.

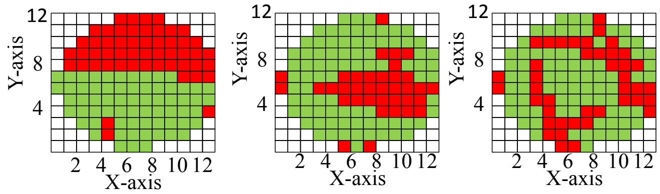


Figure 2. The spatial failure patterns are shown for three outlier wafers corresponding to the single-wafer clusters.

IV. CONCLUSIONS

In this paper, we develop a novel methodology to extract systematic spatial failure patterns across wafers to help improve the product yield of circuit designs. The objective is to extract these systematic failure patterns in the presence of significant random variation and defects, and wafers with abnormal spatial signatures. First, features of the spatial signatures are extracted using SVD. Second, complete-link hierarchical clustering is performed on the reconstructed signatures. Finally, a binary dictionary is learnt to further filter random effects from the systematic spatial failure patterns. The efficacy of the proposed method has been demonstrated by an industrial data set where the systematic spatial failure patterns were extracted in the presence of random effects, while detecting outlier wafers.

REFERENCES

[1] Semiconductor Industry Associate, *International Technology Roadmap for Semiconductors*, 2013.

[2] E. Yilmaz, S. Ozev, O. Sinanoglu and P. Maxwell, “Adaptive Testing: Conquering Process Variations,” *IEEE ETS*, pp. 1-6, 2012.

[3] W. Zhang, X. Li, F. Liu, E. Acar, R. Rutenbar and R. Blanton, "Virtual Probe: A Statistical Framework for Low-Cost Silicon Characterization of Nanoscale Integrated Circuits," *IEEE Trans. on Computer-Aided Design*, vol. 30, no. 12, pp. 1814-1827, 2011.

[4] W. Zhang, X. Li, E. Acar, F. Liu and R. Rutenbar, “Multi-Wafer Virtual Probe: Minimum-Cost Variation Characterization by Exploring Wafer-to-Wafer Correlation,” *IEEE ICCAD*, pp. 47-54, 2010.

[5] X. Li, R. Rutenbar and R. Blanton, “Virtual Probe: A Statistically Optimal Framework for Minimum-Cost Silicon Characterization of Nanoscale Integrated Circuits,” *IEEE ICCAD*, pp. 433-440, 2009.

[6] F. Adly, P. D. Yoo, S. Muhaidat, Y. Al-Hammadi, L. Uihyoung and M. Ismail, “Randomized General Regression Network for Identification of Defect Patterns in Semiconductor Wafer Maps,” *IEEE Trans. on Semiconductor Manufacturing*, vol. 28, no. 2, pp. 143-152, 2015.

[7] N. Kupp, M. Slamani and Y. Makris, “Correlating Inline Data with Final Test Outcomes in Analog/RF Devices,” *IEEE DATE*, pp. 1-6, 2011.

[8] A. Strojwas, “Conquering Process Variability: A Key Enabler for Profitable Manufacturing in Advanced Technology Node,” *IEEE International Symposium on Semiconductor Manufacturing*, pp. xxiii-xxxii, 2006.

[9] J. McNames, B. Moon, B. Whitefield and D. Abercombie, “Robust Linear Regression for Modeling Systematic Spatial Wafer Variation”, *Proceedings of SPIE* , pp.87-98, 2005.

[10] K. Huang, N.Kupp, J. M. Carulli and Y. Makris, “Process Monitoring through Wafer-level Spatial Variation Decomposition”, *Interntional Test Conference*, pp. 1-10, 2013.

[11] W. Zhang, X. Li, S. Saxena, A. Strojwas and R. Rutenbar, “Automatic Clustering of Wafer Spatial Signatures,” *IEEE DAC*, pp. 1-6, 2013.

[12] W. Zhang, K. Balakrishnan, X. Li, D. Boning and R. Rutenbar, “Toward Efficient Spatial Variation Decomposition via Sparse Regression,” *IEEE ICCAD*, pp. 162-169, 2011.

[13] W. Zhang, K. Balakrishnan, X. Li, D. Boning, S. Saxena, A. Strojwas and R. Rutenbar, “Efficient Spatial Pattern Analysis for Variation Decomposition via Robust Sparse Regression,” *IEEE Trans. on Computer-Aided Design*, vol. 32, no. 7, pp. 1072-1085, Jul. 2013.

[14] M. Orshansky, S. Nassif and D. Boning, *Design for Manufacturability and Statistical Design: A Constructive Approach*, Springer, 2007.

[15] A. Jain and R. Dubes, *Algorithms for Clustering Data*, PrenticeHall, 1988.

[16] S. Salvador and P. Chan, “Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms,” *International Conference on Tools with AI*, pp. 576-584, 2004.

[17] F. Raab, “Binary Data Compression by Linear Trasformation,” *Computer Graphics and Image Processing*, vol. 2, no. 1, pp. 12-21, 1973.