Variation and Failure Characterization Through Pattern Classification of Test Data From Multiple Test Stages

Chun-Kai Hsu*, Peter Sarson†, Gregor Schatzberger†, Friedrich Leisenberger†,
John Carulli‡, Siddhartha Siddhartha‡, and Kwang-Ting Cheng*
*University of California, Santa Barbara, CA 93106
†ams AG, 8141 Premstaetten, Austria
‡GlobalFoundries, Malta, NY 12020

***Abstract*—We describe a framework for characterizing systematic variations and failures through exploring the hidden patterns of test data from multiple test stages. The framework provides prediction of process variations with a fine resolution based on a limited number of probed process parameters. An unsupervised biclustering technique is then utilized to extract grayscale and binary spatial patterns from process parameters and production test results, respectively, through analyzing both item-to-item and die-to-die correlations in subsets of the test data. A template matching technique exploits these spatial patterns to discover connections between process variations and failures detected by production tests. The proposed framework has been verified by an industrial test dataset of a non-volatile memory product. The discovery of comprehensible correlations between process parameters and some production test items was confirmed by the engineers who have insights to the test dataset.**

I. Introduction

Efficient and effective testing and diagnosis could significantly improve manufacturing yield for complex designs. At different manufacturing stages, different test strategies are used to target different failures and variations. For example, to evaluate the quality and stability of the manufacturing process, wafer acceptance test (WAT, also known as wafer electrical test, WET, or e-test) is performed at the end of processing a wafer on process control monitors that are small devices located on the scribe lines. On the other hand, production test is conducted for each integrated circuit at a later stage to ensure the correctness and quality of every shipped product. As their objectives and test vehicles are different, the relationship and their implication between the test data from different test stages are not clear. As a result, analysis of test data has usually been restricted to data from one test stage only.

The same variations and/or defects can affect performances and characteristics measured at multiple test stages. For instance, variation to a process parameter can affect both Iddq test in the production test stage and the gate-oxide quality test in WAT. Therefore, characterizing systematic variations and failures based on test data from multiple test stages can potentially further improve process control monitoring, yield estimation, outlier wafer/die detection, and failure diagnosis. However, it is challenging to identify the connection between a bad die's syndromes derived from wafer probe test, especially for those caused by systematic failures, and its process parameters.

Representing the production test data in the form of color-coded wafer maps, two-dimensional spatial patterns, formed by either statistics of good/faulty dies or normalized measurement data, are considered as input to analysis of systematic failures. If similar spatial patterns (with respect to shape, size, and location) in wafer maps derived from test data are observed, there is a high probability that the corresponding faulty dies may experience similar systematic variations. Moreover, if the wafer maps derived from a process parameter and from a production test item exhibit a similar spatial pattern, then it should be strong evidence that the failures are strongly related to variations of a certain process step.

For the objective of characterizing systematic variations and failures between two different test stages, there are several challenges:

a) For the production test items with a binary (pass or fail) outcome, finding the correlations between such nonparametric test items and process parameters is a nontrivial task. Compared to parametric test items, nonparametric test items compress details into a single-bit test result, pass or fail, and thus the test data has little information available for further analysis. That is, all failures have the same test syndrome, a fail symbol, but could be caused by different or the same type of variations and defects.
b) For WAT, there are limited number (e.g., five or nine) of sites within a wafer for measuring process parameters, while production tests are conducted for every die on the wafer. Furthermore, sites and dies are in different scales; a site usually covers multiple dies. Thus, data analytics between process parameters and production test measurements is limited by the resolution of the WAT data. Predicting process parameters for every die location in a complete wafer map based on the WAT data from the limited sites can help increase the accuracy for the temporal (i.e., inter-test-stage) analysis.
c) A correlation model hardly fits the test data with wafer-to-wafer and lot-to-lot variations, which are commonly seen in most products. A model tends to overfits test data that come from a single wafer or from a single lot. On the other hand, a model could underfit test data if data from all accessible wafers are used. In other words, a task of fitting a robust and flexible model for test data within a long time period is nontrivial.
d) Random variation and abnormal wafers degrade the accuracy of test data characterization, especially for the WAT data that have relatively few measurements for a wafer. Spatial patterns that are identified in more wafers or for more process parameters are more trustworthy than others that occur in fewer wafers of for fewer process parameters. Any proposed methods should be insensitive to outliers.

In this paper, we propose a framework with several learning and statistical techniques to address these challenges. To overcome challenge a), we invoke pattern classification, which is well studied in the areas of both semiconductor manufacturing and computer vision, to build the connection between parametric process parameters with nonparametric production tests. Two-dimensional wafer maps could reveal some characteristics of systematic failures that binary vectors

‡This paper contains no production test data from GlobalFoundries.

cannot due to inclusion of extra coordinate information. One example is to separate random defects from the test syndromes with systematic failures that follow certain two-dimensional spatial patterns.

To address challenge b), we extend Virtual Probe (VP) [1] to predict process parameters in every die location within a complete wafer map, based on WAT data at limit site locations. VP, based on *compressed sensing*, is a spatial characteristic method that is a proven technique to achieve high accuracy of modeling spatial patterns using a small number of samples. We improve VP for predicting process parameters in every location within a complete wafer map in the die scale.

We then address challenge c) using a *biclustering* technique. A biclustering technique classifies a two-dimensional test dataset into several biclusters. In each bicluster, the dies are similar to each other on the test items and vice versa. Hence, fitting a model using a subset of data (i.e., fitting based on the data in a bicluster) is more efficient and effective than using the entire dataset. Our experimental results show that some spatial patterns can particularly be recognized in a bicluster. Several biclustering algorithms have been published, such as FABIA [2]. As an advantage, FABIA also takes care of challenge d) by the algorithm itself excluding outliers from the identified biclusters. In other words, a random defect will not be classified as a systematic pattern.

There have been some previous studies about discovering correlations between WAT and production wafer sort test (WS). In [3], the authors predicted WAT measurements by WS measurements using regression analysis tools. In [4], the authors estimated the production yield by WAT data through Bayesian model fusion. Since WAT and WS measurements have different spatial resolution, these studies compressed WS data to either fewer data points or a single yield number and hence reduce the resulting accuracy and precision.

Several techniques for characterizing variations on the spatial domain have been proposed. The method proposed in [1], [5]–[7] constructed models of spatial variations based on generalized least square fitting, compressive sensing, Gaussian process, and Bayesian model fusion, respectively. The method of [8]–[10] decomposed wafer-level and die-level variations into spatial patterns. However, these studies did not address the connection between process parameters and spatial variations.

In addition, several studies focused on recognition and classification for wafer map failure patterns were summarized in [11]. For example, the methods proposed in [12], [13] cluster similar defect spatial patterns using unsupervised-learning neural networks and nearest neighbors, respectively. The approach reported in [14] targeted pattern classification problems on large-scale data sets. In [15], the authors exploited pattern mining methodologies for inter-wafer abnormality analysis.

We organize this paper as follows. We define patterns of test data and review the background of VP and FABIA in Section II. In Section III, we detail the implementation of our proposed framework, including spatial modeling with a greater resolution, biclustering, and pattern classification. Then, we demonstrate the efficacy of our proposed methods using industrial data in Section IV. Finally, we conclude in Section V.

II. Background

In this section, we first introduce several categories of patterns that can be extracted from semiconductor test data. Next, we summarize the VP and FABIA algorithms that are integrated into our proposed framework.

A. Pattern Definition

Patterns are the components with discernible regularity hidden in a series of test measurements. Patterns are usually formed due to correlations caused by systematic variations and/or failures. As summarized in [16], test data correlations can be classified into three categories: spatial correlations, inter-test-item correlations, and temporal correlations.

Spatial correlations, also known as die-to-die correlations, indicate that the measurement of a die is somehow correlated to the measurements of the other dies on the same wafer. Spatial correlations are visually observed as unique spatial patterns with respect to the amplitude of measurements of dies on a wafer map. In addition, pass/fail outcomes (i.e., thresholded performance measurements) form a binary pattern, which is a special type of spatial patterns.

Correlations also exist among different test items when, for example, the same test applied multiple times under different electrical or environmental settings, or different tests targeting the same functionality of a chip. Such correlated measurements result in numerical and abstract patterns that can be identified by several techniques, from simple linear regression to complex support vector machines. Moreover, a set of dies with strong inter-test-item correlations may be spatially clustered and form visually interpretable spatial patterns.

Temporal correlations describe the variations across wafers and lots, i.e., measuring the same performance or process parameter at different times. Monitoring temporal correlations reveals the stability and robustness of the manufacturing process. Taking into account temporal correlations can possibly further improve the accuracy and scalability of applications which solely rely on spatial or inter-test-item correlations.

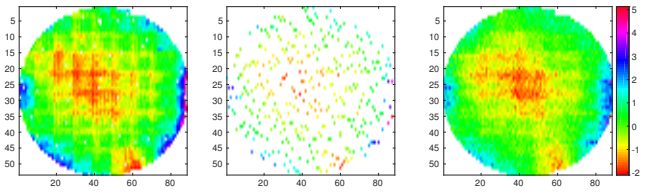
Several studies reported that utilizing correlations in multiple categories improve the quality of test data analytics. For example, in [17], [18], the authors exploited both spatial and inter-test-item correlations for test time reduction and silicon characterization. The method proposed in [19] improves the modeling technique for spatial variations by conducting wafer-to-wafer temporal correlations.

B. Compressed Sensing and VP

Compressed sensing is a signal processing technique for recovering a signal/image using a minimum number of samples. Virtual probe (VP) extends the idea for test data analysis. The essence of VP is to test only a subset of dies at selected locations on a wafer, transform the measurements into spatial frequency domain, and use a statistical algorithm to accurately predict the test measurements of the remaining dies where no measurements were explicitly made [1], [17], [20]. Fig. 1 shows the concept of applying VP to a test item of an industrial product. In this example, the spatial model constructed from 10% randomly sampled dies predict the test values of the remaining 90% dies on the same wafer with sufficiently high accuracy.

The mathematical background of VP is briefly introduced as the following. Let $\{g(x,y);\ x = 1,2,\ldots,P,\ y = 1,2,\ldots,Q\}$ be a performance metric of the die at coordinate (x,y) on a wafer of size $P \times Q$. The spatial variations of $g(x,y)$ can be represented by a two-dimensional linear transform in the frequency domain. In VP, the discrete cosine transform (DCT) is chosen for the transform. Let $\{G(u,v);\ u = 1,2,\ldots,P,\quad v = 1,2,\ldots,Q\}$ be the DCT coefficients after the transform, i.e., the coefficients of different frequencies in the spatial pattern.

The purpose of VP is to accurately recover $g(x,y)$ from a small number, M, of dies at the locations $\{(x_m,y_m;\ m = 1,2,\ldots,M\}$,



(a) Measured values (b) Sample values (c) Predicted values

Fig. 1: (a) Measured values, (b) sampled values, and (c) VP-predicted values of a test item from an industrial product. The values are color-coded to produce the wafer maps.

where $M \ll PQ$. Toward this goal, the following underdetermined linear equation is formulated:

$$A \cdot \eta = B \tag{1}$$

where

$$A = \begin{bmatrix} A_{1,1,1} & A_{1,1,2} & \cdots & A_{1,P,Q} \\ A_{2,1,1} & A_{2,1,2} & \cdots & A_{2,P,Q} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M,1,1} & A_{M,1,2} & \cdots & A_{M,P,Q} \end{bmatrix} \tag{2}$$

$$A_{m,u,v} = \alpha_u \cdot \beta_v \cdot \cos \frac{\pi(2x_m - 1)(u-1)}{2P} \cdot \cos \frac{\pi(2y_m - 1)(v-1)}{2Q} \tag{3}$$

$$\eta = [G(1,1) \ \cdots \ G(P,Q)]^T \tag{4}$$

$$B = [g(x_1, y_1) \ \cdots \ g(x_M, y_M)]^T. \tag{5}$$

Once η is determined by solving Equation (1), the metric values $g(x, y)$ can be recovered by the inverse discrete cosine transform (IDCT). VP assumes η to be sparse [1] to obtain a unique solution of η. That is, most of the DCT coefficients are close to zero, though the locations of the zeros are unknown. We can use *maximum posterior estimation* to statistically solve Equation (1) by reformulating it to

$$\begin{aligned} \underset{\eta}{\text{minimize}} \quad & \|\eta\|_1 \\ \text{subject to} \quad & A \cdot \eta = B \end{aligned} \tag{6}$$

where $\|\eta\|_1$ stands for the L_1-norm of η. Equation (6) can be solved efficiently with linear programming [1].

The generated sparse solution finds the sparsest set of coefficients in the frequency domain that most accurately represent the spatial pattern of the sampled dies. The sampled dies, however, are only a small portion of all the dies on a wafer. Therefore the spatial pattern reconstructed from the sampled dies may not be sufficient if the measurement data exhibit a more random distribution. In other words, if the assumption of sparsity is not valid for the ground truth of a certain test item, finding a sparse solution will not be sufficient to recover the spatial pattern of the test item.

C. Biclustering and FABIA

Biclustering is widely used in bioinformatics field for extracting knowledge from gene expression measurements [21], and has also been generalized for handling two-dimensional dataset. Using semiconductor test data as an example, each row of a test measurement matrix corresponds to a die sample and each column corresponds to a test item. Performing clustering on columns (i.e., clustering similar test items) has a limitation that test items that are somewhat similar may only be similar on a subset of dies, but not all dies. Similarly, clustering rows (i.e., identifying similar dies) has a limitation that dies that are somewhat similar may be similar only for a subset, not all, of test items. Biclustering addresses these limitations and utilizes simultaneous clustering on the row and column dimensions of the test data matrix.

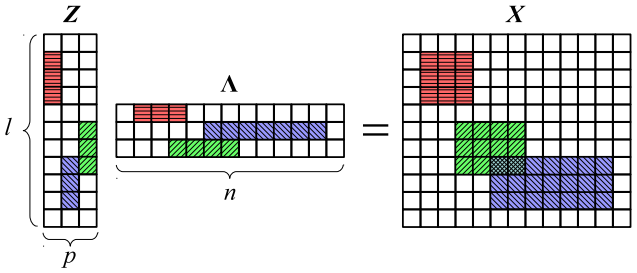


Fig. 2: A multiplicative model with three biclusters. Each rectangular checkerboard represents a matrix. Colored/hatching rectangles are nonzero elements of matrices while white rectangles are zeros. The nonzero elements within a bicluster and within the corresponding prototype and factor vectors have the same color and hatching pattern. Note that the nonzero elements are adjacent to each other for visualization purposes only and additive error $\boldsymbol{\Upsilon}$ is not shown in this figure.

FABIA (Factor Analysis for BIcluster Acquisition) [2] is based on a multiplicative model that can be used to efficiently explore linear dependencies between dies and test items. This paper illustrates how to perform biclustering on test data using FABIA. We briefly summarize the mathematical background of FABIA in the following. The test data are represented as a matrix $\boldsymbol{X} \in \mathbb{R}^{l \times n}$ where l and n are the number of dies and test items, respectively. The element x_{jk} of $\boldsymbol{X}$ corresponds to the measurement of the kth test item in the jth die sample.

In a multiplicative mode, two vectors are similar if one is a multiple of the other, i.e., their correlation coefficient is one (or minus one). Based on this assumption, a bicluster is defined as a pair of a row set and a column set where the rows are similar to each other and the columns are also similar to each other. Such linear dependency is represented by $\boldsymbol{z}\boldsymbol{\lambda}^T$ where $\boldsymbol{\lambda}$ is a prototype column vector with nonzero elements for test items participating in a bicluster, and $\boldsymbol{z}$ is a column vector of factors by which the corresponding prototype columns are scaled. The nonzero elements of $\boldsymbol{z}$ denote the dies that participate in the same bicluster. As shown in Fig. 2, a model with p biclusters is formulated by

$$\boldsymbol{X} = \sum_{i=1}^{p} \boldsymbol{z}_i \boldsymbol{\lambda}_i^T + \boldsymbol{\Upsilon} = \boldsymbol{Z}\boldsymbol{\Lambda} + \boldsymbol{\Upsilon} \tag{7}$$

where $\boldsymbol{\Upsilon} \in \mathbb{R}^{l \times n}$ is additive noise; $\boldsymbol{Z} \in \mathbb{R}^{l \times p}$ and $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times n}$ are the sparse factor matrix and the sparse prototype matrix, respectively. FABIA allows overlapping biclusters, but p should be defined explicitly.

FABIA formulates biclustering as a sparse matrix factorization problem. The measurements of the ith die ($\boldsymbol{x}_i$, the ith row of $\boldsymbol{X}$) can be interpreted by a factor analysis model:

$$\boldsymbol{x}_i = \sum_{j=1}^{p} \boldsymbol{z}_{ij} \boldsymbol{\lambda}_j^T + \boldsymbol{\epsilon}_i = \tilde{\boldsymbol{z}}_i \boldsymbol{\Lambda} + \boldsymbol{\epsilon}_i \tag{8}$$

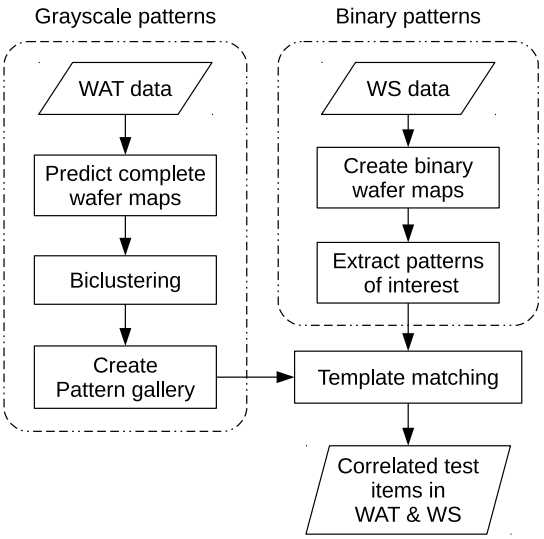


Fig. 3: The application flow of the proposed framework.

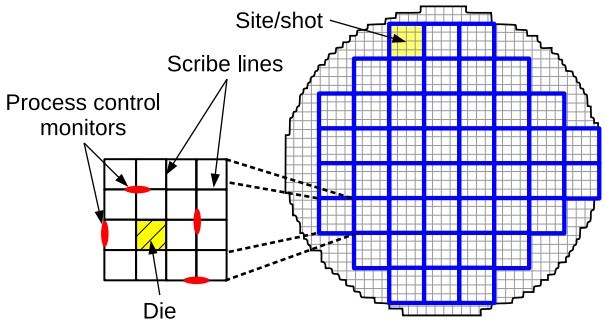


Fig. 4: The spatial relations between sites/shots and dies. Process control monitors are denoted by red ellipses located on the scribe lines. Sites are denoted by rectangles with blue thick edges. Note that the dies close to wafer boundary form incomplete sites that are not emphasized in the figure and are incapable for probing process parameters.

where $\boldsymbol{\epsilon}_i$ is the ith row of the noise matrix $\boldsymbol{\Upsilon}$ and $\tilde{\boldsymbol{z}}_i = (z_{i1}, \ldots, z_{ip})$ denotes the ith row of the factor matrix $\boldsymbol{Z}$. FABIA assumes that each $\tilde{\boldsymbol{z}}_i$ is $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ distributed. The unit covariance matrix indicates that the biclusters are not correlated, and, hence, one bicluster in the data will not be divided into dependent small biclusters. Another assumption is that each $\boldsymbol{\epsilon}_i$ is an independent Gaussian noise and is $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi})$ distributed where $\boldsymbol{\Psi} \in \mathbb{R}^{n \times n}$ is a diagonal covariance matrix.

FABIA then identifies the biclusters (i.e., selecting the model parameters $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ that explain the data best) using *variational expectation maximization* while assuming component-wise independent Laplace priors for both $\tilde{\boldsymbol{z}}$'s and $\boldsymbol{\lambda}$'s. The implementation details are beyond the scope of this paper and can be found in [2].

III. Implementation

A proposed framework for characterizing systematic variations and failures between dataset from different test stages is detailed in this section.

A. Application Flow

Fig. 3 shows the application flow of the proposed framework. The inputs are two different test datasets: one is process control test data, such as WAT data, which usually have relatively few samples due to the limited number of probe points on each wafer, and the other is production test data, such as WS data whose sample count is equal to the number of tested dies. Both datasets include multiple measurements (measured by a set of test items) for each sample. In addition, it is not required that the two datasets are from the same set of wafers although there should exist stronger correlations if they are. Based on these data, the method reports the potential correlated process parameters for the targeted production test items.

Two different methods are applied for extracting patterns from the two types of test data mentioned above. The first method, illustrated in the left part of Fig. 3, is designed to extract grayscale patterns from WAT data. A grayscale pattern is defined as a wafer map in which each element is represented by a real number. VP, described in Section II-B, is employed to predict a complete wafer map with a greater resolution based on the WAT data that only have limited samples for each wafer. Next, FABIA, introduced in Section II-C, is employed to discover biclusters that reveal spatial patterns hidden in the VP-extrapolated WAT data. Such patterns are collected in a *pattern gallery* for further processing.

The second method, shown in the right part of Fig. 3, is designed to extract binary patterns from the production test data, such as the WS data. Binary wafer maps, in which each element has a value of either zero or one indicating a pass or a fail, are derived from thresholded WS data based on predefined specifications. Next, binary patterns are extracted from the wafer map through clustering and classification methods. A template matching technique is then performed to compare the patterns of interest, derived from the WS data, with each pattern in the pattern gallery, extracted from the WAT data, to explore any potential correlations. The following subsections will detail each step of the framework.

B. Spatial Modeling With a Greater Resolution

As shown in Fig. 4, each site (also known as a shot) is a rectangle area, which is usually of the same size as a lithographic photomask, composed of a number of identical dies. The process control monitors are a set of simple circuits located on the scribe lines between dies. The same set of monitors is duplicated for each site because of the exposure using the same photomask. Therefore, the measured process parameters at the process control monitors exhibit the process characteristics of a site, not just of a single die.

The VP algorithm described in II-B is utilized to model wafer-level spatial variations. Fig. 5 shows examples of different strategies for performing VP based on WAT measurements at five sites to predict a complete wafer map. First, Fig. 5(b) shows the predicted wafer map using five samples each of which has a size of a site as shown in Fig. 5(a). The predicted wafer map has a low resolution and becomes pixelated due to the relatively large area of a site. In addition, the predicted values alter sharply from site to site. Second, if each measurement is used only to represent a single die at the center of the corresponding site, as shown in Fig. 5(c), VP derives a predicted wafer map that contains inaccurate high frequency patterns shown in Fig. 5(d) because of an insufficient number of samples. The third strategy results in the best prediction through duplicating each WAT measurement for multiple die locations in a round shape within the site as shown in Fig. 5(e). Fig. 5(f) shows a predicted wafer map that includes more details with a greater resolution.

As the number of samples is very limited, VP (or any other spatial modeling techniques) cannot guarantee the accuracy of the predicted wafer maps. To maximize the accuracy, it is better that

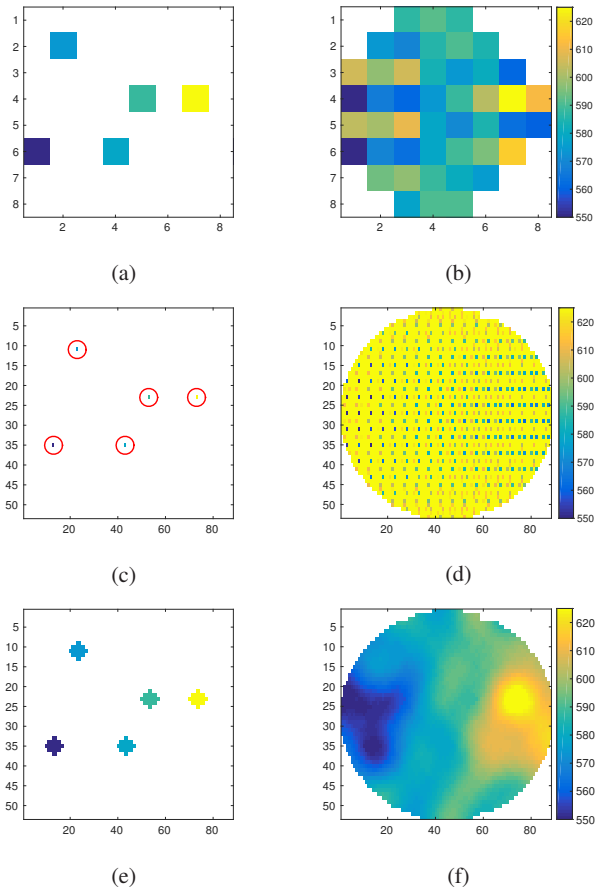


Fig. 5: Color-coded predicted wafer maps of WAT using different sample strategies. (a), (c), and (e) show the samples, as a single site, a single die, and multiple dies, respectively, used by VP for prediction/extrapolation. (b), (d), and (f) show the corresponding predicted wafer maps.

every measurement is used for training. However, there is no ground truth for verifying the accuracy of the predicted WAT wafer maps derived by VP as there are no probed measurements beyond those locations which are already used for VP prediction. On the other hand, cross-validation technique is not valid either due to the lack of probed WAT measurements for an extra independent validation data set.

Fortunately, systematic variations and failures usually influence multiple wafers for most products in mass production. Therefore, if there exist strong correlations between process parameters and test item measurements that are influenced by systematic variations, they should be observed repeatedly in multiple wafers and lots, and thus the reliance on high accuracy of the extrapolated WAT wafer map should be significantly reduced. Hence, as long as a subset of (not necessary all) predicted wafer maps are reasonable accurate, out method should be able to reveal such systematic behaviors.

C. Pattern Extraction Using Biclustering

FABIA is the biclustering method used for extracting spatial patterns from multiple wafer maps. FABIA has two major inputs: a data matrix $\boldsymbol{X}$ and a limit p on the number of biclusters to be identified. Although p has a strong influence on the computation time (basically a cubic time complexity), a slightly large number is the best

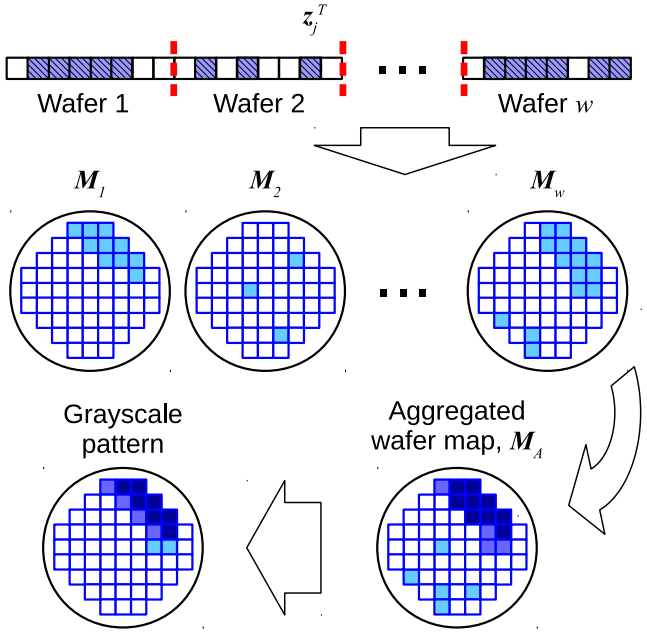


Fig. 6: The procedure of extracting grayscale patterns from a bicluster that is derived by FABIA using WAT data.

since FABIA will only identify as many biclusters as necessary to explain the data. The contents of $\boldsymbol{X}$ are dependent on the types of patterns for which we target to extract.

1) Grayscale Patterns in Process Parameters: For a total of w wafers with n WAT test items used in analysis, we have wn predicted wafer maps derived from VP. An $l \times n$ matrix $\boldsymbol{X}$ is created using the predicted WAT measurements where $l = \sum_{i=1}^{w} l_i$ is the total number of dies and l_i is the number of dies in the ith wafer. A row of $\boldsymbol{X}$ consists of n WAT measurements of one die, and a column of $\boldsymbol{X}$ records measurements of one WAT item. Note that the WAT measurements are normalized with respect to each column. FABIA is then preformed on $\boldsymbol{X}$ to find biclusters that will be represented by sparse matrices $\boldsymbol{Z}$ and $\boldsymbol{\Lambda}$.

Fig. 6 illustrates the procedure of extracting spatial patterns from the jth bicluster through analyzing $\boldsymbol{z}_j$, which is the jth column of $\boldsymbol{Z}$. Samples (dies) with larger absolute values in $\boldsymbol{z}_j$ indicate their more significant roles in the jth bicluster. Those factors in $\boldsymbol{z}_j$ that represent dies from the kth wafer can be grouped and shown as a wafer map $\boldsymbol{M}_k$ (for example, $\boldsymbol{M}_1$ consists of $z_{1j}, z_{2j}, \cdots, z_{l_1 j}$ in $\boldsymbol{z}_j$). Some of these wafer maps expose similar spatial patterns (e.g., $\boldsymbol{M}_1$ and $\boldsymbol{M}_w$ in Fig. 6) while some wafers do not (e.g., $\boldsymbol{M}_2$). Note that each blank rectangle in $\boldsymbol{z}_j$ and, in turn, in $\boldsymbol{M}$s shown in Fig. 6 has a zero value or a value smaller than a given threshold.

In practice, such wafer maps consist of errors, resulting from WAT measurement errors, modeling errors by FABIA, and prediction errors by VP. To enhance the accuracy of spatial patterns revealed by a bicluster and reduce the impact caused by these errors, we overlap all wafer maps resulting from $\boldsymbol{z}_j$ and derive the cumulative sum for each die location resulting in $\boldsymbol{M}_A = \sum_{i=1}^{w} \boldsymbol{M}_i$ as shown in Fig. 6. This aggregated wafer map, $\boldsymbol{M}_A$, can be treated as an image in grayscale and is then normalized to a range $[0, 1]$. Applying an appropriate threshold (e.g., 0.8) to eliminate errors from various sources results in the final grayscale pattern as shown in the lower-left wafer map of Fig. 6.

Based on the factorization model defined in Section II-C, the

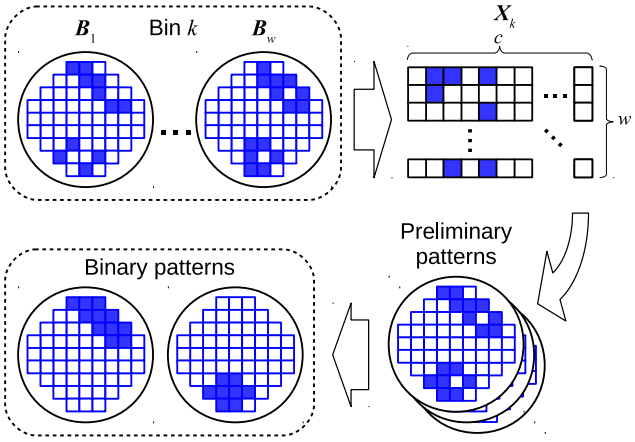


Fig. 7: The procedure of extracting binary patterns based on the binning results.

dies classified into the same bicluster are similar to each other on a subset of WAT items, i.e., these dies have linear inter-test-item correlations. In other words, the measurements of such correlated WAT items increase/decrease simultaneously. Moreover, the intensity of each pixel in a grayscale pattern is proportional to the number of dies that belong to the same bicluster at the location where each pixel is. FABIA does not take spatial locations of dies into account when processing the WAT data. Therefore, if the correlations observed in a bicluster exhibit some spatial pattern such as the grayscale pattern shown in Fig. 6 instead of randomly spreading dots, it indicates a strong connection between a local systematic variation and the corresponding process parameters. We derive grayscale patterns from each bicluster and create a pattern gallery from all identified biclusters for further analysis.

2) Binary Patterns in Production Tests: For production test data, we focus on the pass/fail decisions based on the predefined specifications instead of actual measurements. Hence, the test signature of a die is a n-bit vector where n is the number of test items. For integrated circuit fabrication, classifying dies based on such test signatures is called binnig, i.e., a group of dies that fail on the same set of test items are assigned an identical bin number. The locations of the dies with the same bin number may form spatial patterns that suggest potential systematic variations and failures, especially when the patterns are correlated with some process parameters.

To extract patterns from a target bin (e.g., bin k), the procedures are illustrated in Fig. 7. Assume that there are w wafers used for analysis and the binning results (i.e., the locations of classified dies) are represented by binary wafer maps, $\boldsymbol{B}_1, \boldsymbol{B}_2, \cdots, \boldsymbol{B}_w$. Each die location of $\boldsymbol{B}$s is either zero (not a member of bin k, a blank rectangle) or one (a member of bin k, a solid rectangle). Such wafer maps of bin k can be represented by a $w \times c$ matrix $\boldsymbol{X}_k$ where c is the number of die locations of a wafer map. The ones in ith row of $\boldsymbol{X}_k$ denote the dies in ith wafer that are the members of bin k. The ones in a column of $\boldsymbol{X}_k$ denote the dies in different wafers but at the same die location (one out of c locations) with respect to a wafer map.

In the previous subsection, we have demonstrated that biclustering technique is effective for discovering clusters in a two-dimensional dataset with noise. Therefore, we apply FABIA on $\boldsymbol{X}_k$ to identify any patterns that exist in multiple wafers, i.e., a bicluster (which is a pair of a wafer set and a location set) in $\boldsymbol{X}_k$. Different from the case of processing WAT data, the factorized matrix $\boldsymbol{\Lambda}_k$ of $\boldsymbol{X}_k$ explicitly indicates the location of a pattern by its nonzero elements. As illustrated in the lower-right wafer maps of Fig. 7, each row of $\boldsymbol{\Lambda}_k$ represents a wafer map with preliminary patterns.

Next, a sequence of image processing techniques are performed to enhance and isolate these preliminary patterns: *a*) dilating the image to delineate the outline of the objects of interest, *b*) filling interior gaps for solid objects, *c*) smoothing the objects to compensate the dilating executed before, and then *d*) detecting connected components while excluding abnormal large and tiny objects. These steps result in the final binary patterns as shown in the lower-left of Fig. 7.

D. Template Matching Through Normalized Cross-Correlation

The next step of exploring correlations between WS data and WAT data is formulated as a template matching problem and is solved by normalized cross-correlation (NCC) [22]–[24]. Cross correlation is studied in the area of computer vision to solve the problem of determining the location of a given pattern/template within an image based on the squared Euclidean distance measure. For better identifying similarity, NCC normalizes the image and the template and has the following advantages: *a*) invariant to the image energy (i.e., sum of squares of pixel intensity) under the window containing the template, *b*) having the range $[-1, 1]$ of similarity that is independent of the size of the template, and *c*) invariant to changes in image amplitude, such as brightness or contrast of the image.

The NCC similarity γ of a template t shifted to location (u, v) with respect to an image f is defined as

$$\gamma_{t,f}(u,v) = \frac{\sum_{x,y}\left[f(x,y) - \bar{f}_{u,v}\right]\left[t(x-u,y-v) - \bar{t}\right]}{\sqrt{\sum_{x,y}\left[f(x,y) - \bar{f}_{u,v}\right]^2 \sum_{x,y}\left[t(x-u,y-v) - \bar{t}\right]^2}} \quad (9)$$

where $\bar{f}_{u,v}$ denotes the mean value of pixels of $f(x,y)$ within the region under the shifted t, and $\bar{t}$ denotes the mean value of all pixels in t.

Referring to the application flow in Section III-A, a binary pattern and a grayscale pattern serve as the template t and the image f for Equation (9), respectively. Usually, NCC is performed for each possible offset and the offset $(u_{\max}, v_{\max})$ resulting in the maximum similarity $\gamma_{\max}$ indicates the upper-left corner of the best matching region in the image for the given template. In our framework, templates (WS binary patterns) and images (WAT grayscale patterns) are both of a size of a wafer map. Therefore, we define the maximum similarity of a WS pattern t based on a WAT pattern gallery $\boldsymbol{F}$ as

$$\gamma_{t,\boldsymbol{F}} = \max(\{\gamma_{t,f_i}(0,0) : i = 1, 2, \ldots, |\boldsymbol{F}|, f_i \in \boldsymbol{F}\}) \quad (10)$$

where $\gamma_{t,f_i}(0,0)$ denotes the similarity between t and the ith pattern of $\boldsymbol{F}$ with no shift for t, and $|\boldsymbol{F}|$ denotes the number of patterns in the gallery.

IV. Experimental Results

We evaluate the proposed framework using a test dataset from a non-volatile memory product using 200mm wafers processed on a industry standard 0.35μm mixed-signal technology from ams AG. The dataset includes WAT data and WS data for 300+ wafers with 3500+ dies per wafer. The WAT data have 124 process parameters that were probed from five out of 40 sites per wafer, and the WS data have 95 production test items that are measured for every die.

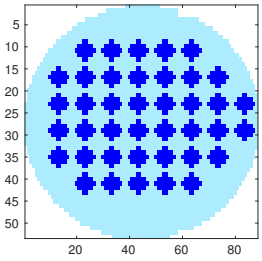


Fig. 8: The mapping between site locations and die locations. Each WAT sample is replaced with 24 die samples in the same region.

A. WAT Data Analysis

As shown in Fig. 4, the WAT data consist of measurements from the process control monitors in five different sites of a wafer, and different sites were selected for different wafers. Before applying VP to derive complete wafer maps, we replaced each WAT sample with 24 die samples having identical measurements that are duplicated from the WAT sample. The locations of these 24 samples are within the region of the site where the WAT sample was probed. The mapping between site locations and die locations are shown in Fig. 8. Each solid round in Fig. 8 indicates 24 dies that are used to replace a site in the same region. As a result, VP has 120 samples per wafer as its input, instead of five samples per wafer in the original WAT data. An example of a predicted WAT wafer map is illustrated in Fig. 5(f). Even though this process parameter is measured at only five sites as shown in Fig. 5(e), the predicted wafer map clearly exposes an upward trend from left to right.

In the next step, we explored biclusters in the predicted WAT data using the FABIA algorithm [2], [25]. The VP-predicted WAT measurements formed a two-dimensional matrix $\boldsymbol{X}$ with 1.1M samples (dies) in rows and 124 WAT items in columns. The limit on the number of biclusters to be identified p was set to 12. Based on Equation (7), these settings resulted in a model with 11 biclusters. The factorized matrix $\boldsymbol{\Lambda}$ are illustrated in Fig. 9. The diagram shows that an item set of a bicluster consists of four to 31 WAT items and the item sets of two different biclusters may have overlap. In addition, each item set has unique WAT items except the item set of Bicluster 11, which is a subset of Bicluster 4's. However, Bicluster 4 is not a subset of Bicluster 11 due to having a different set of dies from each other according to the factorized matrix $\boldsymbol{Z}$. The diagram also indicates that around one-fourth of items (30 out of 124) are linearly independent of any other items and are not clustered into any biclusters.

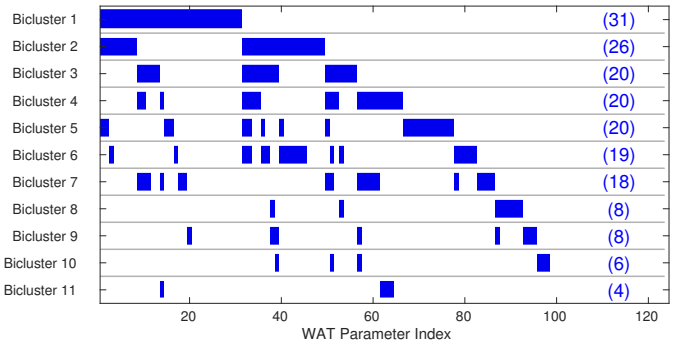
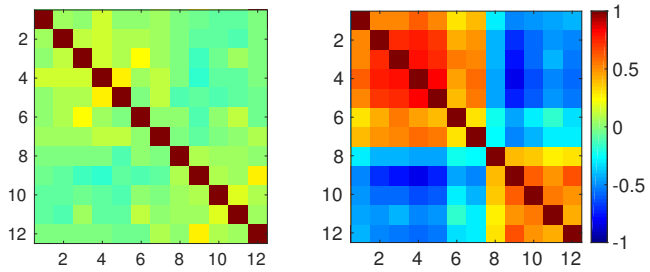


Fig. 9: A diagram showing the factorized matrix $\boldsymbol{\Lambda}$. The solid bars denote nonzero elements of $\boldsymbol{\Lambda}$, i.e., the corresponding WAT items of which a bicluster is composed. The number of WAT items in the item set of a bicluster is given in parenthesis. Note that both bicluster axis and item axis in this diagram are reordered for better visualization.



(a) (b)

Fig. 10: Linear correlation matrix of twelve WAT items of Bicluster 1. Each rectangle represents a color-coded correlation coefficient of two WAT items. (a) and (b) were derived based on S_1' and S_1, respectively.

As mentioned in Section I, systematic variations and failures may not occur globally and only affect a portion of dies in a wafer. That is, with a high probability, such variations appear in a small region of a wafer. Based on the same factorization model described above, assuming that S_1 denotes the die set of Bicluster 1, and S_1' denotes the complement of S_1, i.e., the die set that consists of dies not in S_1. S_1 and S_1' include 23% and 77% of the total number of dies, respectively. Fig. 10 shows the color-coded correlation matrices of twelve WAT items in Bicluster 1 for these two die sets. Every correlation coefficient shown in Fig. 10(a) except those on the diagonal is quite small and is in the range from -0.16 to 0.27. Hence, we can conclude no (or very weak) linear correlations among these WAT items based on S_1'. On the other hand, the correlations shown in Fig. 10(b) are significantly stronger (-0.82 is the strongest coefficient) and become evident for the existence of local systematic variations in S_1.

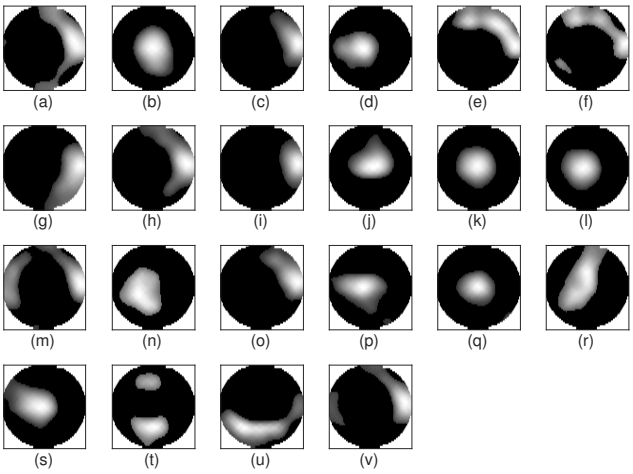
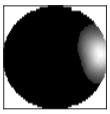
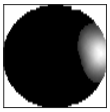
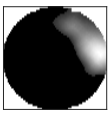
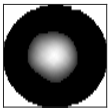


Fig. 11: A grayscale pattern gallery. The intensity of each wafer map is expressed within a range from zero (black) to one (white). Note that the regions outside of a wafer map should be in black (zero values) and are shown as white regions for better visualization.

TABLE I: Matches Between WS Items and WAT Parameters

Match	WS Item(s) Name	Pattern	WAT Parameters Name	Pattern	Similarity γ	Comprehensible Correlations*
1	Volts1stErase, TailErase/ProgramFunc, TailFunctionProg, BasicFunctional, EnduranceRead, RetentionRead/ICell		CWET, CW2, CW2I, CW1, CW1MEF, TWC, JET1, QPMZ1, QPMZ2, QPMZ3		0.78	High
2	SATrip_ICellD0, SATrip_ICellD1, ..., SATrip_ICellD15		CWET, CW2, CW2I, CW1, CW1MEF, TWC, JET1, QPMZ1, QPMZ2, QPMZ3		0.75	Median
3	Volts1stProgram		TMFBL1, WU1, WU2		0.58	Low
4	Volts1stProgram		CWHPY1, CWHPY2, HBNNB, JET2, LQ1, LQ2, UDP1, UPD2		0.53	High
5	MarginHighIRef, MarginLowIRef		SEJG1, SEJG2, SEJG3, XFGG1, XFGG2		0.47	None

*Confirmed by ams AG.

Fig. 12: Binary pattern examples. Note that the regions outside of a wafer map should be in black (zero values) and are shown as gray regions for better visualization.

B. Matches Between Process Parameters and Product Tests

Based on the procedure described in Section III-C1, grayscale patterns of WAT parameters were extracted from the factorized matrix $\boldsymbol{Z}$, which indicates the die set of each bicluster. Referring to Equation (8), a part of dies in a die set positively contribute to $\boldsymbol{X}$ while the rest of the dies contributing negatively. Therefore, each aggregated wafer map from $\boldsymbol{Z}$ revealed two patterns (one representing the positive part and the other representing the negative part of a bicluster) and formed a pattern gallery with 22 grayscale patterns as shown in Fig. 11. Even though some patterns are similar, such as Fig. 11(k) and Fig. 11(l), they were extracted from different biclusters with distinct test item sets and represent different correlations. These spatial patterns abstracted from 124 WAT parameters illustrate some potential systematic variations.

On the other hand, there are 95 test items in the WS data and 23 of them are nonparametric test items with only pass/fail outcomes. 95 test items are classified into 76 test groups and the test items in the same group target similar circuit functions. Based on the procedure described in Section III-C2, binary patterns were extracted from the test results of each test group (i.e., the union of the faulty dies detected by each test item in a test group). Fig. 12 illustrates some binary pattern examples.

For several critical WS items that are related to the quality of memory circuits, the template matching technique, NCC, was performed to examine the most possible candidates in the grayscale pattern gallery. Table I lists five matches that were discovered in this industrial test dataset and were sorted by the NCC similarities. Each row in the Table I indicates the WS item set with the corresponding binary pattern, the WAT parameter set with the corresponding grayscale pattern, and the similarity between these two patterns of a discovered match. The last column reports the level of existing comprehensible correlations between the WAT parameters and the WS items, which are confirmed by the process engineer of ams AG.

Each Match shows a pattern for a WS item set and matches with the WAT pattern derived from the listed parameters. A direct correlation between the WS items and each single WAT parameter is not obvious for the majority of cases and cannot be easily identified. Most of the time several WAT parameters have to shift together to result in a WS fail. The block of the chips related to the WS items has to be analyzed taking the shifts of multiple WAT parameters into account to get a better understanding of a possible correlation or in the best case to get the confirmation of the correlation. However, the proposed framework successfully reveals such correlations.

The correlation between WAT parameters and WS items for Match 1 and Match 2 is comprehensible for the WAT parameters `CWET`, `CW1`, and `JET1` that correspond to the same element. This element is frequently used in the block of the chip responsible for the WS items listed. WAT parameters `QPMZ1` and `QPMZ3` might also contribute to the correlation although the contribution is not as obvious as for the other parameters mentioned above. The WS `SATrip_ICellD#` items can be influenced by a diode related to the WAT parameter `CW2`. The diode is used as a sensing element. If the leakage of the diode shifts, the corresponding WS items of Match 2 will also show a shift. A possible contribution to the correlation of all other WAT parameters is not directly comprehensible.

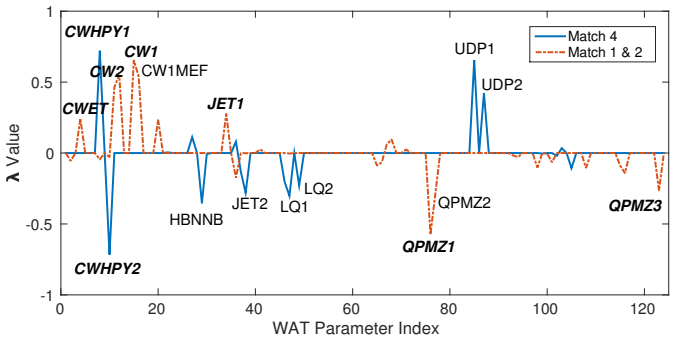


Fig. 13: A plot showing the significance of WAT parameters in Match 1, Match 2, and Match 4. The parameters that are confirmed having comprehensible correlations with WS items are labeled in bold italic.

The correlation between the WS item `Volt1stProgram` and the listed WAT parameters for Match 4 is comprehensible since the WS item and most of the WAT parameters are related to the memory element used in the circuitry. A correlation is obvious for most of the WAT parameters (especially for `CWHPY1` and `CWHPY2`) except for `JET2`, `LQ1`, and `UDP1`. Match 3 and Match 4 show that the same WS item is matched with different WAT parameters. Match 3 is related to the sensing of the memory element while Match 4 is directly related to the memory element. For Match 5, the similarity is already below 0.5 and the WS pattern is more off-center than the WAT pattern. There is no known possible correlations for Match 5.

The listed WAT parameters in Table I are further analyzed based on their significance in a bicluster. An element with a larger absolute value in $\boldsymbol{\lambda}$ of Equation 7 indicates a test item that plays a more significant role in a bicluster. In Match 1, `CW1` and `QPMZ1` are the two most significant parameters that exhibit the WAT grayscale pattern due to their large absolute λ values as shown in Fig. 13. In Match 2, the parameter `CW2` with comprehensible correlations with WS items also has a relatively large λ. Moreover, `CWHPY1` and `CWHPY2` in Match 4, which are confirmed having comprehensible correlations with the WS `Volt1stProgram` item, are the two most significant parameters that exhibit the grayscale pattern. The interpretation of Fig. 13 and the explanation of Table I are coherent.

V. Conclusions

In this paper, we propose a framework for characterizing systematic variations and failures through exploring the hidden patterns of test data from different test stages. The framework utilizes the spatial patterns extracted from both process parameters with a limited number of probed measurements and production tests with binary outcomes. The proposed framework is performed on an industrial test dataset and has successfully revealed some comprehensible correlations between WAT parameters and WS items. The results of such silicon characterization can be used to discover parametric variations and weak links in the manufacturing process.

VI. Acknowledgments

The UCSB authors acknowledge the support of Semiconductor Research Corporation (SRC). The authors acknowledge ams for providing full production data from several ams products for this study. The authors would also like to thank John Carulli and Siddhartha Siddhartha of GlobalFoundries for stimulating discussion of various aspects of test data analytics.

References

[1] X. Li, R. Rutenbar, and R. Blanton, "Virtual probe: A statistically optimal framework for minimum-cost silicon characterization of nanoscale integrated circuits," in *Proc. IEEE/ACM Int'l Conf. on Computer-Aided Design (ICCAD)*, Nov. 2009.

[2] S. Hochreiter, U. Bodenhofer, M. Heusel, A. Mayr, A. Mitterecker, A. Kasim, T. Khamiakova, S. Van Sanden, D. Lin, W. Talloen *et al.*, "FABIA: factor analysis for bicluster acquisition," *Bioinformatics*, vol. 26, no. 12, pp. 1520–1527, Apr. 2010.

[3] S. Devarakond, J. McCoy, A. Nahar, J. Carulli, S. Bhattacharya, and A. Chatterjee, "Predicting die-level process variations from wafer test data for analog devices: A feasibility study," in *Latin American Test Workshop (LATW)*, Apr. 2013.

[4] A. Ahmadi, H.-G. Stratigopoulos, A. Nahar, B. Orr, M. Pas, and Y. Makris, "Yield forecasting in fab-to-fab production migration based on bayesian model fusion," in *Proc. IEEE/ACM Int'l Conf. on Computer-Aided Design (ICCAD)*, Nov. 2015, pp. 9–14.

[5] F. Liu, "A general framework for spatial correlation modeling in VLSI design," in *Proc. IEEE/ACM Design Automation Conf. (DAC)*, Jun. 2007.

[6] N. Kupp, K. Huang, J. Carulli, and Y. Makris, "Spatial estimation of wafer measurement parameters using gaussian process models," in *Proc. Int'l Test Conf. (ITC)*, Nov. 2012.

[7] S. Zhang, X. Li, R. Blanton, J. Machado da Silva, J. Carulli, and K. Butler, "Bayesian model fusion: Enabling test cost reduction of analog/rf circuits via wafer-level spatial variation modeling," in *Proc. Int'l Test Conf. (ITC)*, Oct. 2014.

[8] S. Reda and S. Nassif, "Accurate spatial estimation and decomposition techniques for variability characterization," *IEEE Trans. on Semiconductor Manufacturing*, vol. 23, no. 3, pp. 345–357, Aug. 2010.

[9] W. Zhang, X. Li, S. Saxena, A. Strojwas, and R. Rutenbar, "Automatic clustering of wafer spatial signatures," in *Proc. IEEE/ACM Design Automation Conf. (DAC)*, May 2013.

[10] K. Huang, N. Kupp, J. Carulli, and Y. Makris, "Process monitoring through wafer-level spatial variation decomposition," in *Proc. Int'l Test Conf. (ITC)*, Sep. 2013.

[11] F. Adly, P. Yoo, S. Muhaidat, and Y. Al-Hammadi, "Machine-learning-based identification of defect patterns in semiconductor wafer maps: An overview and proposal," in *Proc. IEEE Int'l Parallel Distributed Processing Symp. Workshops (IPDPSW)*, May 2014.

[12] F.-L. Chen and S.-F. Liu, "A neural-network approach to recognize defect spatial pattern in semiconductor fabrication," *IEEE Trans. on Semiconductor Manufacturing*, vol. 13, no. 3, pp. 366–373, Aug. 2000.

[13] T. Yuan, W. Kuo, and S. J. Bae, "Detection of spatial defect patterns generated in semiconductor fabrication processes," *IEEE Trans. on Semiconductor Manufacturing*, vol. 24, no. 3, pp. 392–403, Aug. 2011.

[14] M.-J. Wu, J.-S. Jang, and J.-L. Chen, "Wafer map failure pattern recognition and similarity ranking for large-scale data sets," *IEEE Trans. on Semiconductor Manufacturing*, vol. 28, no. 1, pp. 1–12, Feb. 2015.

[15] N. Sumikawa, L. C. Wang, and M. S. Abadir, "A pattern mining framework for inter-wafer abnormality analysis," in *Proc. Int'l Test Conf. (ITC)*, Sep. 2013.

[16] F. Lin, C.-K. Hsu, and K.-T. Cheng, "Learning from production test data: Correlation exploration and feature engineering," in *IEEE Asian Test Symp. (ATS)*, Nov. 2014, pp. 236–241.

[17] C.-K. Hsu, F. Lin, K.-T. Cheng, W. Zhang, X. Li, J. M. Carulli Jr., and K. M. Butler, "Test data analytics — exploring spatial and test-item correlations in production test data," in *Proc. Int'l Test Conf. (ITC)*, Sep. 2013.

[18] S. Zhang, F. Lin, C.-K. Hsu, K.-T. Cheng, and H. Wang, "Joint virtual probe: Joint exploration of multiple test items' spatial patterns for efficient silicon characterization and test prediction," in *Proc. Conf. Design, Automation, and Test in Europe (DATE)*, Mar. 2014.

[19] 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," in *Proc. IEEE/ACM Int'l Conf. on Computer-Aided Design (ICCAD)*, Nov. 2010, pp. 47–54.

[20] H.-M. Chang, K.-T. Cheng, W. Zhang, X. Li, and K. Butler, "Test cost reduction through performance prediction using virtual probe," in *Proc. Int'l Test Conf. (ITC)*, Sep. 2011.

[21] S. C. Madeira and A. L. Oliveira, "Biclustering algorithms for biological data analysis: A survey," *IEEE/ACM Trans. on Computational Biology and Bioinformatics*, vol. 1, no. 1, pp. 24–45, Jan. 2004.

[22] J. Lewis, "Fast normalized cross-correlation," in *Vision interface*, vol. 10, no. 1, 1995, pp. 120–123.

[23] K. Briechle and U. D. Hanebeck, "Template matching using fast normalized cross correlation," in *Proc. SPIE Optical Pattern Recognition XII*, vol. 4387, Mar. 2001, pp. 95–102.

[24] S.-D. Wei and S.-H. Lai, "Fast template matching based on normalized cross correlation with adaptive multilevel winner update," *IEEE Trans. on VLSI Image Processing*, vol. 17, no. 11, pp. 2227–2235, Nov. 2008.

[25] J. K. Gupta, S. Singh, and N. K. Verma, "MTBA: Matlab toolbox for biclustering analysis," in *IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions*, Jul. 2013, pp. 94–97.