A Graph-Theoretic Approach for Spatial Filtering and Its Impact on Mixed-Type Spatial Pattern Recognition in Wafer Bin Maps

Ahmed Aziz Ezzat, *Member, IEEE*, Sheng Liu, Dorit S. Hochbaum, and Yu Ding, *Senior Member, IEEE*

***Abstract*—Statistical quality control in semiconductor manufacturing hinges on effective diagnostics of wafer bin maps, wherein a key challenge is to detect how defective chips tend to spatially cluster on a wafer—a problem known as *spatial pattern recognition*. Recently, there has been a growing interest in *mixed-type* spatial pattern recognition—when multiple defect patterns, of different shapes, co-exist on the same wafer. Mixed-type spatial pattern recognition entails two central tasks: (1) spatial filtering, to distinguish systematic patterns from random noises; and (2) spatial clustering, to group filtered patterns into distinct defect types. Observing that spatial filtering is instrumental to high-quality mixed-type pattern recognition, we propose to use a graph-theoretic method, called adjacency-clustering, which leverages spatial dependence among adjacent defective chips to effectively filter the raw wafer maps. Tested on real-world data and compared against a state-of-the-art approach, our proposed method achieves at least 46% gain in terms of internal cluster validation quality (i.e., validation without external class labels), and about 5% gain in terms of Normalized Mutual Information—an external cluster validation metric based on external class labels. Interestingly, the margin of improvement appears to be a function of the pattern complexity, with larger gains achieved for more complex-shaped patterns.**

***Index Terms*—Clustering, Graph theory, Pattern recognition, Spatial data science, Unsupervised learning, Wafer bin maps.**

I. Introduction

INTEGRATED circuits (IC), colloquially known as chips, are essential to most, if not all, electronic devices. The central step in IC manufacturing is wafer fabrication, in which a batch of chips are fabricated on round-shaped silicon wafers through a series of complex electrochemical processes including slicing silicon-rich ingots into round-shaped thin wafers, wafer oxidation and material deposition, photolithography, ion implantation, and etching [1]. Once fabricated, all wafers undergo an operational quality performance test, known as wafer probing, in which chips are labeled as functional or defective by passing an input signal and collecting the corresponding output. This step results in a two-dimensional graphical representation called a *wafer bin map*—a gridded representation of a wafer in which each grid point represents the spatial location of a chip and is assigned a binary value (e.g., 0 or 1) denoting a functional or a defective chip, respectively. Figure 1 shows examples of three different wafer bin maps resulting from wafer probing tests, where defective chips are denoted by red squares.

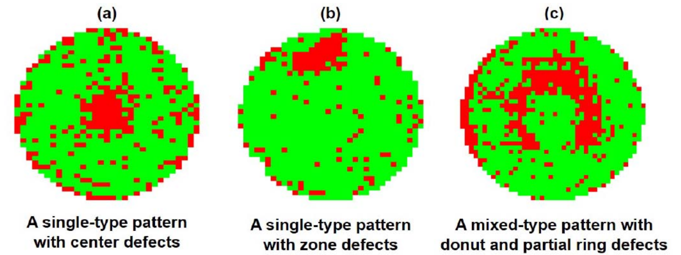

Fig. 1. Examples of wafer bin maps. Panels (a-b) are *single-type* patterns, i.e., one defect pattern per wafer, while Panel (c) shows a *mixed-type* defect pattern. *Random patterns* due to inherent process variation are represented by the scattered red squares on the wafer maps, which overlap with *systematic patterns* (e.g., center, zone, donut) that are attributed to assignable root causes.

A careful analysis of wafer bin maps is pivotal to quality control efforts in the semiconductor manufacturing industry. By investigating the spatial defect patterns on the fabricated wafers, i.e., how the defective chips tend to spatially cluster, one can infer instrumental insights about the root causes of defect occurrence, and subsequently suggest corrective actions to mitigate future failures. This problem, often referred to in the literature as *spatial pattern recognition* (SPR), is the focus of this paper. SPR is of extreme importance to pinpoint possible root causes of failures in wafer fabrication. In fact, several spatial defect patterns in wafer bin maps can be directly traced to common root causes of failure. For instance, a circular-shaped, center-located defect pattern, as shown in Figure 1(a), often corresponds to chemical stains or mechanical equipment faults [2], [3], while a zone-shaped, edge-located defect pattern, as shown in Figure 1(b), can be traced to uneven polishing or edge-die effects [4]. A center-located, donut-shaped

Manuscript received October 24, 2020; revised January 5, 2021; accepted February 16, 2021. Date of publication March 2, 2021; date of current version May 5, 2021. The work of Dorit S. Hochbaum was supported in part by NSF Award under Grant CMMI-1760102. *(Corresponding author: Ahmed Aziz Ezzat.)*

Ahmed Aziz Ezzat is with the Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, 08854, USA (e-mail: aziz.ezzat@rutgers.edu).

Sheng Liu is with the Rotman School of Management, University of Toronto, Toronto, ON M5S 3E6, Canada.

Dorit S. Hochbaum is with the Department of Industrial Engineering and Operations Research, University of California at Berkeley, Berkeley, CA 94720 USA.

Yu Ding is with the Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX 77843 USA.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSM.2021.3062943.

Digital Object Identifier 10.1109/TSM.2021.3062943

0894-6507 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.

defect pattern, as shown in Figure 1(c), is routinely observed in wafer data due to possible tooling problems [5]. These spatial patterns like center, zone, or donut, wherein defective chips "cluster" to form distinct shapes are referred to as *systematic patterns* since they correspond to an assignable root cause. In contrast, randomly scattered defective chips in Figure 1(a)-(c) are called *random patterns*, or *noises*, since they are merely artifacts of random process variation.

In the SPR literature, defect patterns in wafer bin maps can either be *single-type* or *mixed-type*. The former refers to wafer maps that host only one defect pattern, while the latter refers to wafer maps in which two or more defect patterns co-exist. Figure 1(a-b) show examples of single-type defect patterns, whereas Figure 1(c) depicts a mixed-type defect pattern. With the ever-growing increase in scale and sophistication of the wafer fabrication process, mixed-type patterns are increasingly observed in production data. Barring few recent efforts [3], [6]–[10], the problem of mixed-type SPR has received less attention relative to its single-type counterpart.

A typical SPR analysis of a wafer bin map, be it hosting single- or mixed-type patterns, involves two pillar tasks: (1) Spatial filtering, i.e., to de-noise raw wafer data by separating systematic from random patterns; and (2) Spatial clustering, i.e., to group the filtered patterns into one or more sub-clusters pertaining to different types of defect patterns (e.g., center, donut). The overwhelming majority of the literature has been devoted to improving the effectiveness of the second task, namely spatial clustering, among which those that are based on model- or density-based clustering [2], [11]–[14], kernel-based clustering [15], [16], similarity-based metrics such as correlograms, nearest-neighbor measures, or Voronoi-based partitioning [17]–[19], feature extraction-based approaches such as those utilizing Hugh transforms, single value decomposition, or mask-based features [20]–[23], decision trees and manifold learning [24], [25], regression-based consensus networks and ensemble learning [24], [26], [27], and neural networks, especially those based on adaptive resonance theory [28]–[33], or on deep learning-based architectures [34]–[42].

On the other hand, methods for the first task, namely spatial filtering, are mostly dominated by *ad hoc* heuristics intended to pre-process or denoise the raw wafer data, with an implicit assumption that the deficiencies of a poorly designed filtering step will be ultimately corrected in the second task (spatial clustering). While this assumption may be acceptable for single-type SPR, we claim that spatial filtering is of extreme importance to mixed-type SPR. Motivated by a similar observation, an algorithm called *connected path filtering* (CPF) has been recently proposed to filter mixed-type wafer bin maps [3]. The authors proposed to pair CPF with a spatial clustering model that acts on the filtered data to produce the final SPR results. CPF is a heuristic algorithm that searches all possible connected paths of defective chips on a wafer and only keeps those paths that are longer than a pre-set threshold, M.

While valuable on its own, our analysis of multiple wafer maps, as we will elaborate in the sequel, has revealed two main limitations of the CPF approach. First, CPF does not directly leverage local spatial neighborhood information, but instead, it disregards all defective chips that do not belong to a connected path which is longer than a globally pre-set value, M. In other words, if a chip is labeled as "functional," while all of its neighbors are not, CPF does not make use of the local neighborhood information to possibly re-assign the label of this chip. A direct consequence of this limitation is that the filtered results may end up having irregular shapes (few functional chips surrounded by large groups of defective chips, or vice-versa), which may severely compromise the quality of the downstream clustering task. To demonstrate this first limitation, let us take a look at Figure 2, where Panel (a) shows a raw wafer bin map with a mixed-type pattern comprising partial ring and donut defects. The results from CPF (Panel b) clearly show an irregular shape due to either functional chips for which the values should have been updated to defective based on their local neighborhood, or alternatively, defective chips which should have been re-labeled as functional. This irregularity in the shapes misleads the downstream spatial clustering (performed using a mixture model—to be reviewed later) to mistakenly identify some defective chips as independent sub-clusters. The results in Panel (c) appear to be more visually appealing with a clear visual distinction between the two overlapping defect types, making the downstream clustering method (performed using the same mixture model) relatively straightforward. The result in Panel (c) is in fact, produced by our proposed filtering approach, for which the details are unraveled in Section II.

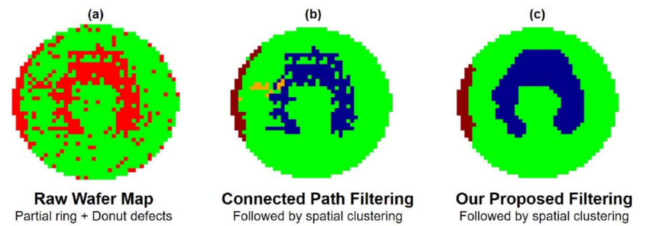

Fig. 2. Analysis of a wafer map with overlapped partial ring and donut defects. (a): raw wafer bin map, (b): CPF with $M = 5$, and (c): Better filtering results produced by our proposed method.

The second limitation of the CPF approach is its choice of M. Our findings, to be presented in Section III, suggest that there does not appear to be a default value for M that works universally well for all wafers and combinations of defect types, making the choice of M wafer-specific. As outlined by the authors in [3], this choice is made via interaction with domain experts. This limitation may severely hamper the applicability of CPF in practice. Given the large production volumes typical of wafer production lines, the need for domain experts to constantly weigh in and update the value of M can be extraordinarily inefficient and thus impractical.

Motivated by the need to address those two limitations, we propose an innovative approach for mixed-typed SPR in wafer bin maps. Our major finding is that a graph-theoretic approach which leverages the local spatial dependence structure can considerably improve the spatial filtering step, and ultimately, the overall SPR quality. Specifically, we propose to use the adjacency-clustering (AC) method, which was originally

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.

introduced by [43] for yield prediction. Although technically similar, the main function of AC in our work is different from that in [43]: we focus on extracting systematic defect patterns (i.e., diagnostic analysis), rather than yield prediction (i.e., prognostic analysis). This distinction drives a fundamental departure in what constitutes a cluster: In [43], they define a cluster as a group of chips with homogeneous yield level, while our approach defines a cluster based on its chips' membership in either the set of systematic or random defect pattern clusters. AC is closely tied to Markov Random Field (MRF) models which have been successfully applied in image segmentation and restoration [44], [45], spatial clustering [46], and wafer thickness variation analysis [47].

We couple the proposed spatial filtering method with a mixture model for spatial clustering. Based on the numerical experiments with real-world data, we show that our approach outperforms the state-of-the-art method in [3] by at least 46% in terms of internal cluster validation quality (i.e., validation without external information about class labels), and about 5% in terms of Normalized Mutual Information—an external cluster validation metric which makes use of externally provided class labels. The mixture model used for spatial clustering is the same as that used by [3] so that the improvements in the final clustering quality are solely attributed to our proposed filtering method. Interestingly, the margin of improvement appears to be a function of the defect pattern complexity, with larger gains achieved for more complex-shaped patterns.

In summary, our main contribution is to propose a graph-theoretic spatial filtering method which effectively distinguishes the random noises that are prevalent in wafer bin maps from systematic patterns, which are often attributed to well-studied root causes. There is an overwhelming evidence in the distant and recent literature that a poorly designed filtering method can have substantial detrimental impacts on the SPR performance in wafer bin maps, even with the emergence of powerful deep learning-based approaches [6], [36], [38], [40]. Unlike existing filtering and pre-processing methods, our approach fully leverages the spatial dependence information (via its graph-theoretic structure), and is not vulnerable to arbitrary parameter selections (e.g., filter size or length thresholds) which can be wafer-specific, thus requiring continuous intervention of domain experts. Furthermore, our method is shown to have a desirable combinatorial structure which can be solved in polynomial time by a minimum-cut algorithm. When coupled with a spatial clustering approach (in this paper, a mixture model), substantial improvements in SPR performance are detected, owing to the abovementioned advantages. In principle, we could replace this mixture model by any clustering or classification approach, thus emphasizing the generality and substantial benefit brought about by our proposed spatial filtering method.

We conclude this Section by describing the organization of this paper. In Section II, we elucidate the building blocks of our proposed approach, which comprises the details of the AC approach to filter the wafer maps, coupled with an advanced mixture model to further group the AC-filtered results into one or more sub-clusters corresponding to different systematic defect patterns. Section III presents our case study which details the analysis of twelve real-world wafer bin maps exhibiting complex multi-type defect patterns. Section IV concludes this paper and highlights future research directions.

II. Our Approach

We represent a wafer map with n chips by $(d_1, d_2, \ldots, d_n)$, where $d_i \in \mathbb{N}$ is the number of defects on chip i. The locations of chips can be modeled as a graph $G = (V, E)$ where nodes denote chips and the edges define the neighborhood relationship, i.e., we have an edge $[i, j] \in E$ when chip i and chip j are adjacent to each other on the wafer map. According to the neighborhood system, each chip can have at most four neighbors (rook-move neighborhood) or eight neighbors (king-move neighborhood).

Our SPR framework consists of two stages: a spatial filtering stage, and a spatial clustering stage, both of which are clustering tasks, yet they serve different purposes. In the first stage, namely spatial filtering, AC partitions the wafer map into two clusters, such that one of them only includes those chips that form systematic defect patterns. As a result, we are able to separate the systematic patterns (those caused by assignable root causes) from the random patterns (artifacts of random process variation). In the second stage, the AC-filtered results are further partitioned into one or more sub-clusters using a mixture model called the infinite warped mixture model (iWMM). Each sub-cluster corresponds to a type of systematic defect pattern, e.g., a center or zone, as shown in Figure 1.

A. Adjacency-Clustering for Spatial Filtering

We sketch here the AC model from [43] and present how it can be adapted into the mixed-type SPR problem. The AC model aims to partition the set of chips into clusters such that chips belonging to the same cluster behave similarly and tend to be adjacent to each other. This AC concept is motivated by the spatial dependence among adjacent chips, which aligns well with the concept of the systematic defect patterns on a wafer map where defective chips tend to spatially cluster. In the case of binary defect data (i.e., $d_i \in \{0, 1\}$), the AC model will find two clusters: the first cluster corresponds to the set of systematic defect patterns, while the other corresponds to random defect patterns.

The clustering decisions are cluster labels x_i for $i \in V$. Chips with the same label form a single cluster. The objective function of AC includes a deviation cost function and a separation cost function. The deviation cost function measures how x_i deviates from the observed value d_i, while the separation cost function captures the difference in assigned labels of adjacent chips. Let $f_i(x_i, d_i)$ denote the deviation cost functions associated with node $i \in V$ and $g_{ij}(x_i - x_j)$ denote the separation cost functions for edge $[i, j] \in E$. AC can be formulated as the following integer program:

$$\min \sum_{i \in V} f_i(x_i, d_i) + \sum_{[i,j] \in E} g_{ij}\big(x_i - x_j\big) \qquad \text{(AC)}$$

$$s.t. \;\; x_i \in X \quad \forall\, i \in V, \qquad (1)$$

where X is the set of allowable labels of each chip. In our application, we have $X = \{0, 1\}$.

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.

The clustering results depend on the trade-off between the two cost functions. When the separation function values are relatively larger than the deviation function values, the resulting clusters will be more contiguous (the spatial smoothing effect is more significant). If the separation costs are too large, the whole wafer map would be forced to have the same label so the separation cost is minimized. On the other hand, when the separation costs are small, the assigned labels will be close to the original observational values and the spatial filtering effect is less notable in the clustering result. The AC model has a statistics foundation from Markov random fields (MRF) wherein solving for x_i is finding the maximum a posterior (MAP) estimate of the degradation model with an MRF prior [45]. Different forms of separation and deviation functions reflect different distributional assumptions of MRF.

The neighborhood system also plays an important role in the AC results. When the rook-move neighborhood structure is assumed, the clustering only looks for defect patterns that grow horizontally or vertically. By contrast, with the king-move structure, the clustering will identify defect patterns that exhibit more complex shapes such as ring and donut patterns. Therefore, the king-move structure can work better for complicated clustering tasks, as those prevalent in mixed-type defect detection (See Figure 2(c) for an example).

When both d_i and x_i are binary ($X = \{0, 1\}$), as in our SPR application, then the AC model reduces to the problem called minimum s-excess [48]:

$$\min \sum_{i \in V} w_i x_i + \sum_{[i,j] \in E} u_{ij} z_{ij}, \tag{AC-BIN}$$

$$s.t. \quad z_{ij} \geq x_i - x_j \ \forall \ [i, j] \in E, \tag{2}$$

$$z_{ij} \geq x_j - x_i \ \forall \ [i, j] \in E, \tag{3}$$

$$x_i \in \{0, 1\} \ \forall \ i \in V, \ \ z_{ij} \in \{0, 1\} \ \forall \ [i, j] \in E. \tag{4}$$

where $z_{ij} = |x_i - x_j| \in \{0, 1\}$ indicates the difference in the label values of chip i and j, while w_i is the deviation cost of chip i, and u_{ij} is the separation cost associated with the pair of chips. More specifically, $w_i = f_i(1, 0) > 0$ for chips with $d_i = 0$ and $w_i = -f_i(0, 1) < 0$ for chips with $d_i = 1$: (1) when $d_i = 0$, we will incur a penalty of $f_i(1, 0)$ for labeling $x_i = 1$ and zero penalty otherwise, so the associated deviation cost is $f_i(1, 0) \cdot x_i$ and $w_i = f_i(1, 0)$; (2) when $d_i = 1$, we will incur a penalty of $f_i(0, 1)$ when assigning $x_i = 0$ and zero penalty otherwise, hence the associated deviation cost is $f_i(0, 1) \cdot (1 - x_i)$; After dropping the constant, we get $w_i = -f_i(0, 1)$. And $u_{ij} = g_{ij}(1) > 0$ for all pairs (the separation cost is 0 if $z_{ij} = 0$). The reduction to the minimum s-excess is attainable for any deviation and separation functions.

The minimum s-excess problem can be solved in polynomial time with a minimum-cut algorithm applied to an appropriately defined graph [48].

Proposition 1: The adjacency-clustering model with binary label values (AC-BIN) can be solved in polynomial time.

Proof: First, we can verify that the constraint matrix of (AC-BIN) is totally unimodular and therefore, (AC-BIN) can be solved in polynomial time. Second, the constraints also correspond to the dual of the minimum cost network flow problem. Then finding the solution to (AC-BIN) is equivalent to finding the minimum cut on a graph adapted from G [48]. ■

The algorithm constructs a graph $G_{st} = (V \cup \{s, t\}, A_{st})$ as follows: First we add to G a source node s and a sink node t, and each edge $[i, j]$ is replaced by two arcs (i, j) and (j, i) with the same capacity of u_{ij}. For each node $i \in V$ with a positive w_i, we add an arc of capacity w_i from the node to the sink. For each node $j \in V$ with a negative weight w_j, we add an arc of capacity $-w_j$ from the source. Then the defective cluster (the set of chips with $x_i = 1$) is the source set of a minimum cut in G_{st}. The computational results indicate that this algorithm can solve instances with thousands of chips within seconds, which facilitates its real-time adoption in practice.

If d_i and x_i take more than two values, the AC model can still be solved in polynomial time for convex deviation and separation functions. Specifically, for "bilinear" separation cost functions (i.e., $g(x_i - x_j) = u_{ij} \cdot (x_i - x_j)$ if $x_i \geq x_j$ and $u_{ji} \cdot (x_j - x_i)$ otherwise) and any convex deviation function, [48] devised an algorithm that solves the problem in the running time of a single minimum cut (and the running time of finding the minima of the convex deviation functions). This time complexity was also shown by [48] to be the best that can be achieved. When the separation cost function is not "bilinear" but convex, the Lagrangian relaxation technique can be applied for the polynomial time algorithm [49].

After solving AC with binary label values, each chip is assigned a new label. The chips with a label value of one form a cluster that contains systematic defect patterns, while the chips with a label value of zero are filtered out. In other words, the original defects recorded on the zero-label chips are treated as random defect patterns of nonassignable causes, to be marked off by the spatial filtering stage and thus no longer deemed defects in the subsequent spatial clustering stage. The spatial filtering result depends on the relative magnitude of the separation costs and deviation costs (the relative differences between w_i and u_{ij} in AC-BIN). As we show in Section III, our numerical analysis suggests that there is a set of parameter values that yields consistently high quality SPR results.

B. Infinite Warping Mixture Model for Spatial Clutering

Given the AC-filtered results, we apply iWMM [50] to group the resulting systematic patterns into sub-clusters pertaining to distinct types of defect patterns. Before we elaborate on the details of iWMM, we first briefly discuss the motivation of using it in our setting. iWMM was first proposed by [50] and then adopted by [3] to spatially cluster the wafer maps that are filtered via CPF. In our approach, we keep the iWMM as our spatial clustering method, because iWMM is a highly potent multi-class clustering method and lends itself well to the SPR problem (more about this in the following). Additionally, by using the same spatial clustering method as that used by [3], we ensure that the improvements in SPR quality are mainly attributed to our proposed filtering method.

The benefit of using iWMM in spatial clustering of wafer defect patterns is two-fold. First, in iWMM, the number of sub-clusters corresponding to the number of defect types on a wafer is estimated rather than specified *a priori*—a common

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.

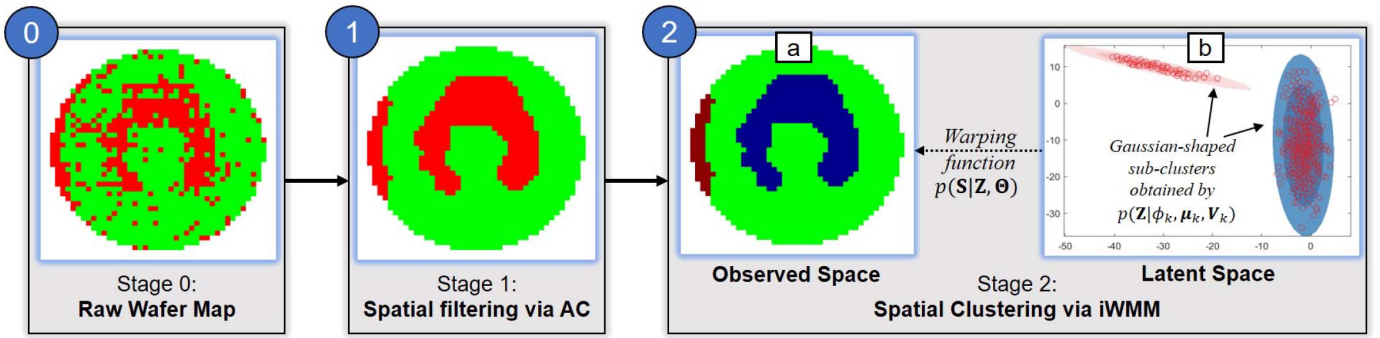

Fig. 3. Our SPR consists of spatial filtering via the AC method (Stage 1) and iWMM for spatial clustering (Stage 2). iWMM assumes the non-Gaussian-shaped sub-clusters in the observed space (2a) are obtained by warping Gaussian-like sub-clusters in the latent space (2b).

shortfall of most clustering methods. Second, and more importantly, defects in wafer maps tend to have non-Gaussian shaped patterns (such as donut and ring). This invalidates the assumptions of many classical model-based clustering approaches that assume the clusters themselves follow a certain parametric distribution (most commonly a Gaussian). The iWMM method relaxes this assumption by making the parametric distribution assumption on the clusters in a latent space, which are related to the original complex-shaped clusters through a non-linear transformation called a warping function. Through this warping, complex non-Gaussian-like shapes in the observed space can be represented by simple Gaussian-like shapes in the latent space. Clustering is then performed in the latent space using a model-based clustering technique (e.g., Gaussian mixture). Figure 3 shows an example of how iWMM works within our framework.

We briefly describe the key details of the iWMM method in our problem setting and interested readers can refer to the Appendix for more details. As evident in Figure 3, two building blocks constitute the essence of iWMM: (1) a warping function to match the observed spatial locations of the AC-filtered results with a set of latent spatial coordinates in the latent space, and (2) a model-based clustering method which determines the clustering assignments in the latent space. The authors in [50] propose to use a Gaussian process latent variable model (GPLVM) [51] as a warping function, and an infinite Gaussian mixture model (iGMM) as a model-based clustering method. We briefly review both in the sequel.

Using the notation from Section II, we denote by $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_n]^T$ the set of spatial locations of the defective chips in the AC-filtered results, i.e., for which $x_i = 1$, where $n = \sum_i x_i$, and $\mathbf{s}_i \in \mathbb{R}^2$. The set $\mathbf{S}$ in the observed space corresponds to a set of latent coordinates in the latent space denoted by $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]^T$, where $\mathbf{z}_i \in \mathbb{R}^2$. The ultimate goal of iWMM is to find a vector of assignments in the latent space, denoted as $\mathbf{A} = [a_1, \ldots, a_n]^T$, where $a_i \in \mathbb{Z}^+$ denotes the membership of the ith chip to a particular sub-cluster.

GPLVM is used to map (or warp) the transformed spatial locations $\mathbf{Z}$, which are assumed to follow a Gaussian distribution, into the observable space where $\mathbf{S}$ can have a non-Gaussian distribution. For GPLVM, the conditional probability of $\mathbf{S}$ given $\mathbf{Z}$ is expressed as in Eq. (5) [51].

$$p(\mathbf{S}|\mathbf{Z}, \mathbf{\Theta}) = (2\pi)^{-n}|\mathbf{\Sigma}|^{-1} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{S}^T\mathbf{\Sigma}^{-1}\mathbf{S}\right)\right), \quad (5)$$

where $\mathrm{tr}(\cdot)$ is the trace function, $|\cdot|$ is the determinant operation, and $\mathbf{\Theta}$ is an $n \times n$ covariance matrix whose ith and jth entry holds the covariance between a pair of observations $\mathbf{z}_i$ and $\mathbf{z}_j$. To determine the entries of $\mathbf{\Theta}$, a stationary parametric covariance function $C(\cdot)$ is selected, which depends on the lag between a pair of inputs through a set of hyperparameters $\mathbf{\Theta}$. A popular choice for $C(\cdot)$ is the so-called squared exponential covariance [52], which is employed in this work.

Once the warping function is established, iGMM is used for spatial clustering in the latent space by assuming that the kth mixture component (or sub-cluster) in the latent space follows a Gaussian distribution with a mean and precision matrix, denoted by $\boldsymbol{\mu}_k$ and $\mathbf{V}_k$, respectively. Each mixture component is associated with a mixture weight, ϕ_k. The mathematical expression for iGMM is presented in Eq. (6).

$$p\left(\mathbf{z}|\phi_k, \boldsymbol{\mu}_k, \mathbf{V}_k\right) = \sum_{k=1}^{\infty} \phi_k \mathcal{N}\left(\mathbf{z}|\boldsymbol{\mu}_k, \mathbf{V}_k^{-1}\right) \quad (6)$$

A detailed procedure to fit the iWMM to a set of observed spatial locations $\mathbf{S}$ is proposed in [50], where the latent coordinates $\mathbf{Z}$, assignments $\mathbf{A}$, as well as remaining parameters are inferred through a Markov Chain Monte Carlo (MCMC)-based procedure. The implementation codes for iWMM have been made publicly available [53] and we use them for our numerical analysis in Section III.

III. Application to Real-World Wafer Map Data

In this section, we evaluate the performance of our proposed SPR approach on real-world wafer bin maps. We then derive key insights about its performance relative to a state-of-the-art approach using widely recognized clustering quality metrics.

A. Data Description

We extract twelve wafer maps from a public dataset that is widely cited in the semiconductor manufacturing community [54], [55]. While the original dataset contains a large number of wafer maps, we select twelve wafers so as to

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.

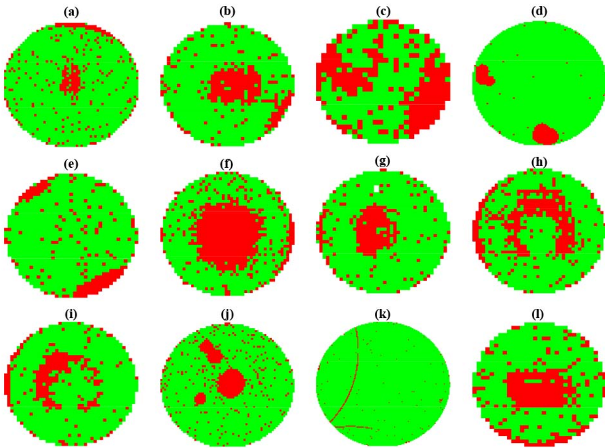

Fig. 4. Twelve wafer maps with mixed-type defect patterns. (a), (f), (g), and (l): wafer maps with center and partial ring defects. (b) and (j): wafer maps with center and zone defects. (c), (d), and (e): wafer maps with two zone defects, (h) and (i): wafer maps with donut and partial ring defects, (k): a wafer map with two disconnected scratch defects.

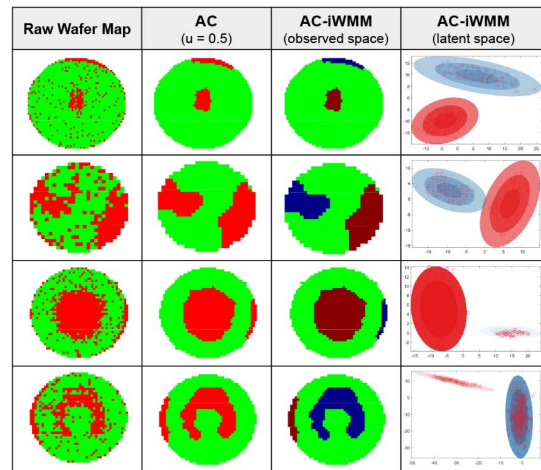

Fig. 5. Results from a subset of wafers, starting from the raw maps (first column), to AC filtering using $u = 0.5$, and then to iWMM clustering, in both the original and latent spaces (third and fourth columns, respectively).

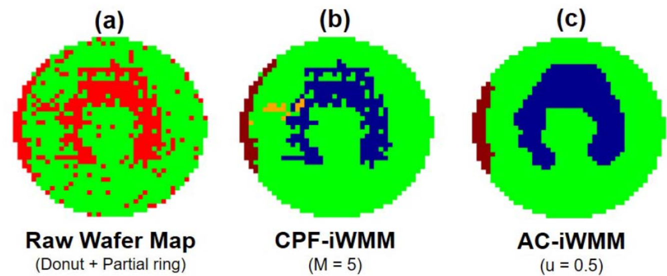

Fig. 6. Visual comparison of CPF-iWMM (b) and AC-iWMM (c) on a wafer with donut and partial ring defects. Unlike AC, CPF fails to separate the two sets of defects, causing iWMM to mistakenly flag a separate sub-cluster.

(1) reflect different mixed-type defect patterns of various complexity, and (2) to resemble as close as possible to the six wafer maps analyzed by [3] (which we do not have access to) in order to provide a fair comparison of our proposed approach relative to the state-of-art approach in the literature. Figure 4 displays the twelve chosen wafer maps, where the red and green squares depict the defective and functional chips, respectively.

B. Results and Discussion

Hereinafter, we denote our proposed approach as AC-iWMM where adjacency-clustering for spatial filtering is coupled with the infinite warping mixture model for spatial clustering. The benchmark in comparison is the state-of-the-art filtering approach in [3], which is denoted hereinafter as CPF-iWMM where the connected path filtering (CPF) algorithm for spatial filtering is followed by iWMM for spatial clustering. Therefore, the fundamental difference between our approach and the benchmark lies in the spatial filtering stage, for which the impact on the quality of SPR is shown to be instrumental.

We test the AC model with a king-move neighborhood system, and standardize $f_i(0, 1) = f_i(1, 0) = 1$ (i.e., $|w_i| = 1$) for all $i \in V$, and set $u_{ij} = u$ for all $[i, j] \in E$. The value of u thus controls the spatial filtering level. In theory, we can choose the value of u through a cross validation procedure as described by [43]. As discussed in Section II-A, having a too large or too small value of u is not ideal for the spatial filtering. We observe that for almost all defect patterns, the choice of $u = 0.5$ achieves a sensible trade-off between the deviation and separation costs, and consistently yields superior performance in filtering various defect pattern combinations. This is in contrast to CPF for which there does not appear to be a value for its main parameter M that works universally well for different defect types (findings to be discussed in the sequel). We implement the CPF algorithm following the description of the method by [3], while for iWMM, we adapt the codes available in [53]. Figure 5 shows the results of the AC-iWMM approach for a subset of the wafers, starting from the raw map, to AC-filtered results, to the clustering results in both the original and latent spaces. The visual results for all wafers are shown in Figure 10 in the Appendix.

1) Visual Comparisons: Before we present the quantitative results, we first draw some insights based on visual comparisons between our approach and the benchmark, CPF-iWMM. Figures 6 and 7 show the results of both approaches on two wafer maps. The first wafer map, illustrated in Figure 6, hosts donut and partial ring defect patterns. By virtue of AC filtering, iWMM is able to distinguish the two types of defects into two separate sub-clusters that are spatially distinct. This is achieved by correctly smoothing out the random noises between the two patterns with the use of local neighborhood information. In contrast, CPF mistakenly identifies some chips located in proximity to both sub-clusters as a separate sub-cluster, as it overlooks local neighborhood information. We note that iWMM was run at the same parameter settings for AC-iWMM and CFP-iWMM so the difference between the two sets of results is solely attributed to the spatial filtering approach.

Another illustrative example is shown in Figure 7, in which the wafer map exhibits two zone defects. Again, CPF fails to separate the two sets of defective chips, causing iWMM to mistakenly flag a new separate sub-cluster. This is in contrast to AC-iWMM which yields a clear distinction between the two

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.

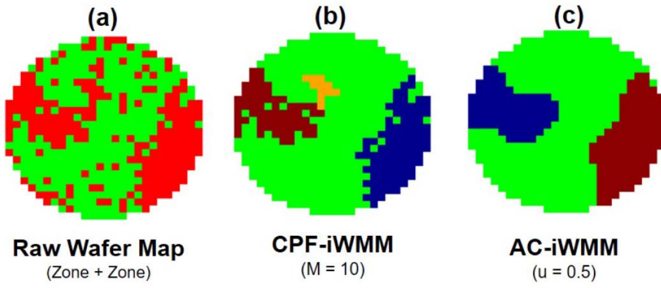

Fig. 7. Visual comparison of CPF-iWMM (b) and AC-iWMM (c) on a wafer with two zone defects. In contrast to AC, CPF fails to separate the two sets of defective chips, causing iWMM to mistakenly flag a separate sub-cluster.

sub-clusters. We note that this problem cannot be alleviated by simply tuning the value of M because the set of chips that are mistakenly flagged by CPF-iWMM are connected to one of the true sub-clusters, and hence, CPF will always treat it as one connected path. AC-iWMM does not keep this set of chips after AC filtering because the neighborhood information is utilized to smooth them out, reducing potential mishaps in the subsequent iWMM clustering stage.

2) Quantitative Comparisons: The clustering results obtained by both approaches are then evaluated using two sets of performance metrics that are known in the SPR literature as *internal* and *external* indices [56]. Assuming $\hat{\mathbf{A}} = [\hat{a}_1, \ldots, \hat{a}_n]^T$ and $\mathbf{A} = [a_1, \ldots, a_n]^T$ are the sets of predicted and true cluster assignments, respectively, internal indices assess SPR quality when the underlying ground truth is not available, that is, without access to the set $\mathbf{A}$. External indices, on the other hand, make use of $\mathbf{A}$ to validate the estimated SPR results.

Let us denote by $\mathbf{G}$ the center of all coordinates in $\mathbf{S}$. Similarly, $\mathbf{G}^k$ denotes the center of the coordinates in $\mathbf{S}^k = \{\mathbf{s}_i^k\}_{i:\hat{a}_i=k}$, i.e., the coordinates of the observations assigned to the kth sub-cluster (for which $\hat{a}_i = k$). Two widely recognized internal indices are the Calinski-Harabasz (CH) index and the Generalized Dunn index. The Calinski-Harabasz (CH) index calculates a weighted ratio of between-cluster and within-cluster dispersion and is defined in Eq. (7) [57]. By definition, a higher value for the CH index indicates a better performance.

$$\mathrm{CH}\Big(\mathbf{S}, \mathbf{S}^1, \ldots, \mathbf{S}^{\hat{K}}\Big) = \frac{n-\hat{K}}{\hat{K}-1} \frac{\sum_{k=1}^{\hat{K}} n_k \left\|\mathbf{G}^k - \mathbf{G}\right\|^2}{\sum_{k=1}^{\hat{K}} \sum_{i:\hat{a}_i=k} \left\|\mathbf{s}_i^k - \mathbf{G}^k\right\|^2}, \tag{7}$$

where $\|\cdot\|$ is the Euclidean norm, and $\hat{K}$ is the predicted number of sub-clusters.

The Generalized Dunn Index (GDI) defines a similar ratio, as expressed in Eq. (8) [58]. A higher value for GDI indicates a better performance.

$$\mathrm{GDI}\Big(\mathbf{S}^1, \ldots, \mathbf{S}^{\hat{K}}\Big) = \frac{\min_{k \neq k'} \frac{1}{n_k + n_{k'}} \Big(\sum_{i:\hat{a}_i=k} \left\|\mathbf{s}_i^k - \mathbf{G}^k\right\| + \sum_{j:\hat{a}_j=k'} \left\|\mathbf{s}_j^{k'} - \mathbf{G}^{k'}\right\|\Big)}{\max_k \max_{i \neq j:\hat{a}_i=\hat{a}_j=k} \left\|\mathbf{s}_i^k - \mathbf{s}_j^k\right\|}. \tag{8}$$

In addition to these internal indices, we test the performance of our approach on a set of widely recognized external indices. The motivation is that, in practice, domain experts can provide the ground truth for a set of testing wafer data which can be used to assess the performance of the competing SPR approaches. Since our dataset does not have the "ground truth," or in other words the set $\mathbf{A}$, we reconstruct the ground truth by applying a pattern reconstruction technique which iterates over every pixel of the raw map and updates its value using a weighted sum of its surrounding pixels to generate an output image [59]. For our application, we used a 3×3 neighborhood system with $\frac{4}{9}$ weight. We also observed that this weight selection had minimal impacts on the final results.

Three prevalent external indices are the Rand index (RI), adjusted Rand index (ARI), and normalized mutual information (NMI). The first two metrics are based on counting pairs of observations on which the predicted clustering results agree or disagree with the true clustering assignment. Specifically, let us assume that K and $\hat{K}$ are the true and predicted number of sub-clusters, respectively, and that n_{ij} denote the number of observations that are common in the ith sub-cluster of $\mathbf{A}$ and the jth sub-cluster of $\hat{\mathbf{A}}$. Now, let us define γ as the number of pairs pertaining to the same sub-cluster in $\mathbf{A}$ and to the same sub-cluster in $\hat{\mathbf{A}}$, while β, on the other hand, denotes the number of pairs pertaining to different sub-clusters in $\mathbf{A}$ and different sub-clusters in $\hat{\mathbf{A}}$. With the above notations, RI, first introduced in [60], can be defined as:

$$\mathrm{RI}\Big(\mathbf{A}, \hat{\mathbf{A}}\Big) = \frac{\gamma + \beta}{\binom{2}{n}} \in [0, 1], \tag{9}$$

where in case of perfect clustering, RI $= 1$, and in general, the higher its value, the better.

The second metric is the adjusted Rand index, or in short ARI, and is computed as follows:

$$\mathrm{ARI}\Big(\mathbf{A}, \hat{\mathbf{A}}\Big) = \frac{\binom{2}{n}(\gamma + \zeta) - \big[(\gamma + \beta)(\gamma + \tau) + (\tau + \zeta)(\beta + \zeta)\big]}{\binom{2}{n}^2 - \big[(\gamma + \beta)(\gamma + \tau) + (\tau + \zeta)(\beta + \zeta)\big]}, \tag{10}$$

where τ denotes the number of pairs pertaining to the same sub-cluster in $\mathbf{A}$ and to different sub-clusters in $\hat{\mathbf{A}}$, while, ζ denotes the number of pairs pertaining to different sub-clusters in $\mathbf{A}$ and to the same sub-cluster in $\hat{\mathbf{A}}$. Similar to RI, a higher value of ARI indicates better performance.

The third external metric is NMI [61], which is an information-theoretic metric that measures the amount of information that $\mathbf{A}$ and $\hat{\mathbf{A}}$ share, and is expressed as in Eq. (12). NMI ranges between 0 and 1, with higher values indicating better performance.

$$NMI\Big(\mathbf{A}, \hat{\mathbf{A}}\Big) = \frac{I\Big(\mathbf{A}, \hat{\mathbf{A}}\Big)}{H\Big(\mathbf{A}, \hat{\mathbf{A}}\Big)} \in [0, 1], \tag{11}$$

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.

TABLE I
INTERNAL INDICES (CH AND GDI) FOR ALL 12 WAFERS. BOLD-FACED VALUES INDICATE BEST PERFORMANCE. * DENOTES $u = 0.40$

	CH (↑)			GDI (↑)		
Wafer #	CPF-iWMM (M=5)	CPF-iWMM (M=10)	AC-iWMM (u=0.50)	CPF-iWMM (M=5)	CPF-iWMM (M=10)	AC-iWMM (u=0.50)
1	539.0	539.0	**703.0**	**.195**	**.195**	**.195**
2	99.0	99.0	**304.4**	.045	.045	**.281**
3	517.4	307.3	**593.4**	**.229**	.176	.223
4	7800.5	7800.5	**8266.1**	**.303**	**.303**	.300
5	1302.5	1214.6	**1565.8**	.210	.213	**.215**
6	85.1	219.3	**254.8**	.160	.285	**.302**
7	185.9	185.9	**222.8**	**.276**	**.276**	**.276**
8	102.3	91.8	**148.4**	.167	.139	**.260**
9	114.2	200.4	**240.8**	.208	.302	**.320**
10	455.8	341.9	**682.3**	.130	.141	**.228**
11*	38.8	**137.9**	54.7	.011	**.226**	.184
12	190.4	176.1	**218.0**	.236	.284	**.285**

such that

$$I\left(\mathbf{A}, \hat{\mathbf{A}}\right) = \sum_{i=1}^{K} \sum_{j=1}^{\hat{K}} \frac{n_{ij}}{n} \log\left(\frac{n_{ij}/n}{\left(\sum_{j=1}^{\hat{K}} n_{ij}\right)\left(\sum_{i=1}^{K} n_{ij}\right)/n^2}\right)$$

$$H\left(\mathbf{A}, \hat{\mathbf{A}}\right) = -\sum_{i=1}^{K} \sum_{j=1}^{\hat{K}} \frac{n_{ij}}{n} \log\left(\frac{n_{ij}/n}{\left(\sum_{i=1}^{K} n_{ij}\right)/n}\right). \tag{12}$$

Tables I and II summarize the comparison results, in terms of internal and external metrics, respectively. We have included results of the CPF approach at two different values of the threshold, namely $M = 5$ and $M = 10$. With a lack of a systematic way to select M, those two values are selected as representatives of a low and high value, respectively. All internal and external metrics are computed using the statistical programming software `R`. Specifically, values of RI and ARI are computed by using functionalities in the library `fossil` [62], while NMI is computed by calling the library `NMI`. All internal indices are computed by using functionalities in the library `clusterCrit` [63].

As shown in Tables I and II, we find that, in all wafers and across all metrics, AC-iWMM either outperforms or comes as a close second relative to CPF-iWMM with $M = 5$ or $M = 10$. We also note that the performance of the CPF approach is, in many cases, sensitive to the choice of M. As a case in point, varying M from 5 to 10 in wafer #6 changes an internal metric like CH by as much as 158%, and an external metric like NMI by up to 3%. More importantly, there is not a choice of M that consistently outperforms the other. For instance, we note that a choice of $M = 5$ for wafer #3 outperforms that of $M = 10$. In contrast, a choice of $M = 10$ for wafer #6 renders consistently better results than $M = 5$ across all metrics. This suggests that the choice of M may be wafer-specific and requires expert judgment (as acknowledged in [3]). As opposed to CPF, the choice of $u = 0.5$ for AC is shown to provide consistently satisfactory performance across all defect combinations, except for the scratch patterns, where a value of $u = 0.4$ worked best. Note that, in principle, CPF is expected to perform considerably well in scratch patterns, since by design, scratch patterns are line defects, which, if longer than a carefully selected threshold (for this wafer, $M = 10$), can be naturally characterized by CPF. Nevertheless, our approach is still able to effectively distinguish the scratch defects as shown in the visual results in Figure 10 (wafer #11). We also note how changing M from 10 to 5 for this wafer results in a substantial deterioration in performance for CPF, as shown in Tables I and II.

Table III presents the percentage improvements of AC-iWMM relative to CPF-iWMM at $M = 5$, 10, for all metrics. On average (last row of Table III), AC-iWMM achieves an average improvement of up to 201% over CPF-iWMM in terms of internal metrics, and up to 6% in terms of external metrics. The difference in scale between the improvements in internal and external indices is attributed to how these metrics are defined in the first place; As described earlier, internal metrics are used to assess the clustering quality *sans* externally provided information about the underlying cluster labels. While external validation metrics are perhaps more interpretable than their internal counterparts, the latter can be extremely useful in practice, since it may be cumbersome for experts to constantly weigh in and provide external information about class labels for all tested wafers. In other words, internal validation provides an automated check point to evaluate the method's performance in real-time. To confirm the considerable improvement brought by AC-iWMM, we perform the Wilcoxon signed ranked test, which is a nonparametric statistical test of hypothesis, with its null hypothesis suggesting that the difference between a pair of samples follows a symmetric zero-centered distribution. The resulting p-values are shown in Table IV, wherein improvements are shown to be statistically significant for all metrics at a significance level of 0.01, except for improvements in GDI which are significant at a 0.1 significance level.

Another interesting observation is the magnitude of improvement realized by AC-iWMM over CPF-iWMM as a function of the defect pattern complexity. Specifically, we note that improvements from AC-iWMM are more pronounced for more complex-shaped defect patterns, and are diminishing as the defect patterns become relatively simpler. For instance, substantial improvements (maximal for some metrics) in Tables I and II come from wafer map #9, hosting donut and partial ring defect patterns. The results for that wafer is shown in Figure 8. Understandably, donut and partial ring defect patterns are, by design, intricate shapes, making the distinction between random and systematic defects a much harder task. This is where the AC approach, through exploiting the local spatial information, can play an instrumental role in improving the quality of the spatial filtering step, and eventually, the downstream clustering and pattern recognition. On the other hand, CPF does not make use of local neighborhood information, which causes the downstream clustering to misidentify several defective chips as an independent sub-cluster.

In contrast, wafer map #4 has a relatively simple mixed-type defect pattern, in which the two zone defects are round shaped and far from each other. Furthermore, the random defects outside the two zones are relatively sparse, which makes the filtering task straightforward. Therefore, both methods

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.

TABLE II
EXTERNAL INDICES (RI, ARI, AND NMI) FOR ALL 12 WAFERS. BOLD-FACED VALUES INDICATE BEST PERFORMANCE. * DENOTES $u = 0.40$

	RI (↑)			ARI (↑)			NMI (↑)		
Wafer #	CPF-iWMM (M=5)	CPF-iWMM (M=10)	AC-iWMM (u=0.50)	CPF-iWMM (M=5)	CPF-iWMM (M=10)	AC-iWMM (u=0.50)	CPF-iWMM (M=5)	CPF-iWMM (M=10)	AC-iWMM (u=0.50)
1	.969	.969	**.985**	.938	.938	**.970**	.877	.877	**.916**
2	.927	.927	**.956**	.849	.849	**.908**	.785	.785	**.845**
3	.831	.830	**.856**	.612	.607	**.661**	.660	.647	**.686**
4	**.993**	**.993**	.990	**.986**	**.986**	.979	**.950**	**.950**	.936
5	.976	.974	**.985**	.952	.948	**.969**	.897	.893	**.917**
6	.900	.901	**.943**	.772	.776	**.869**	.751	.770	**.841**
7	.940	.940	**.967**	.877	.877	**.932**	.817	.817	**.876**
8	.919	.922	**.939**	.837	.842	**.877**	.762	.778	**.812**
9	.872	.871	**.923**	.726	.724	**.830**	.699	.711	**.790**
10	.977	.973	**.980**	.954	.946	**.961**	.888	.880	**.906**
11*	.985	**.987**	.985	.968	**.971**	.968	.938	**.946**	.936
12	.894	.885	**.931**	.775	.759	**.851**	.743	.738	**.799**

TABLE III
IMPROVEMENT (IN PERCENTAGE) OF AC-IWMM OVER CPF-IWMM WITH $M = 5$, 10, FOR ALL METRICS (INTERNAL AND EXTERNAL) ACROSS THE 12 WAFERS. GRAY-COLOURED CELLS DENOTE INSTANCES WHERE PERCENTAGE IMPROVEMENT WAS NEGATIVE

	Internal Indices				External Indices					
	CH		GDI		RI		ARI		NMI	
Wafer #	M = 5	M = 10	M = 5	M = 10	M = 5	M = 10	M = 5	M = 10	M = 5	M = 10
1	30.4%	30.4%	0.00%	0.00%	1.65%	1.65%	3.41%	3.41%	4.45%	4.45%
2	208%	208%	524%	524%	3.13%	3.13%	6.95%	6.95%	7.64%	7.64%
3	14.7%	93.1%	-2.62%	26.7%	3.01%	3.13%	8.01%	8.90%	3.94%	6.03%
4	5.97%	5.97%	-0.99%	-0.99%	-0.30%	-0.30%	-0.71%	-0.71%	-1.47%	-1.47%
5	20.2%	28.9%	2.38%	0.94%	0.92%	1.13%	1.79%	2.22%	2.23%	2.69%
6	199%	16.2%	88.8%	5.96%	4.78%	4.66%	12.6%	12.0%	12.0%	9.22%
7	19.9%	19.9%	-0.32%	-0.32%	2.87%	2.87%	6.27%	6.27%	7.22%	7.22%
8	45.0%	61.7%	55.7%	87.1%	2.18%	1.84%	4.78%	4.16%	6.56%	4.37%
9	111%	20.1%	53.9%	5.96%	5.85%	5.97%	14.3%	14.6%	13.0%	11.1%
10	50.1%	100%	75.2%	61.68%	0.32%	0.77%	0.66%	1.58%	1.95%	2.95%
11	41.1%	-60.29%	1594%	-18.6%	0.00%	-0.20%	0.00%	-0.42%	-0.30%	-1.10%
12	14.5%	23.82%	20.31%	0.20%	4.16%	5.17%	9.77%	12.11%	7.62%	8.31%
Avg.	63.3%	45.6%	201%	58.0%	2.36%	2.49%	5.66%	5.93%	5.40%	5.12%

TABLE IV
WILCOXON SIGNED RANK TEST RESULTS. EACH ENTRY SHOWS THE RESULTING p-VALUE FOR THE CORRESPONDING METRIC

Internal Indices				External Indices					
CH		GDI		RI		ARI		NMI	
M = 5	M = 10	M = 5	M = 10	M = 5	M = 10	M = 5	M = 10	M = 5	M = 10
5×10^{-4}	7×10^{-3}	3×10^{-2}	8×10^{-2}	3×10^{-3}	2×10^{-3}	3×10^{-3}	2×10^{-3}	2×10^{-3}	2×10^{-3}

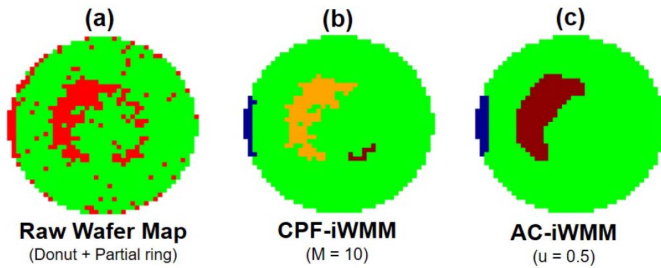

Fig. 8. Visual comparison of AC-iWMM and CPF-iWMM on wafer map #9 with donut and partial ring defects. We note that CPF fails to separate the two sets of defective chips, causing iWMM to mistakenly flag a separate sub-cluster. Here, AC-iWMM achieves substantial improvements over CPF-iWMM owing to its ability to better filter complex-shaped defect patterns.

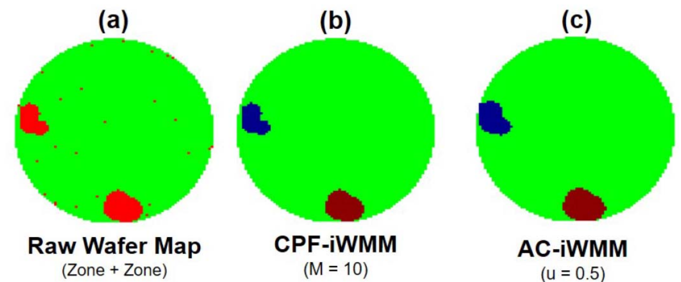

Fig. 9. Visual comparison of AC-iWMM and CPF-iWMM on wafer #4 with two zone defects. We note that both approaches render similar results, visually, and quantitatively. The marginal difference is due to the simplicity of the defect patterns—two zone defects with sparse random noises in the background, which are effectively filtered by both methods.

were able to produce satisfactory performance, with almost negligible visual differences, as shown in Figure 9.

The observations from Figures 8 (for wafer map #9) and 9 (for wafer map #4) validate our conjecture that the difference in performance of AC-iWMM relative to CPF-iWMM hinges on the complexity of the underlying defect patterns. A closer look at the results in Table III suggests the same observation for wafer maps #6 (visual result shown in Figure 10

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.

Wafer #	Raw Wafer Map	AC (u = 0.5)	AC-iWMM (observed space)	AC-iWMM (latent space)
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11*				
12				

Fig. 10. Results from all 12 wafers, starting from wafer id (first column), raw wafer maps (second column), to AC filtering results (third column), and then to iWMM clustering as applied to the AC-filtered data, in both the original and latent spaces (fourth and fifth columns, respectively). * denotes $u = 0.4$.

in the Appendix) and #8 (visual result shown in Figure 6), which have complex-shaped patterns, and hence, the benefit from AC-iWMM appears to be more pronounced. As the wafer fabrication process grows in scale and sophistication, owing to technology upgrades, or an increase in the number of processing steps or the density of chips per wafer, wafers are expected to exhibit more intricate and mixed-type defect patterns. Thus, we anticipate that our proposed approach will generate even more profound impacts.

IV. Conclusion

In this paper we have proposed a spatial pattern recognition framework (AC-iWMM) for detecting mixed-type defect patterns in wafer bin map data—a problem of vital importance to ensuring quality control in the semiconductor manufacturing industry. This framework integrates the adjacency-clustering (AC) model for spatial filtering with an advanced mixture model (iWMM) for spatial clustering. AC has a desirable combinatorial structure and can be solved in polynomial time by a minimum-cut algorithm. By utilizing the local neighborhood information, AC is able to effectively distinguish the systematic patterns from random noises. As a result, iWMM, which subsequently acts on the AC-filtered data, can properly cluster the systematic patterns into different types. We validate the superior performance of AC-iWMM on twelve real-world wafer bin maps exhibiting different mixed-type defect patterns. Based on both visual and quantitative comparisons, AC-iWMM outperforms the state-of-the-art method in the literature, especially for complex-shaped, mixed-type patterns.

The framework proposed herein can be extended in various directions. One interesting idea is to incorporate additional features to aid with mixed-type SPR such as the number of defects per bin on the premise that defects due to the same assignable cause may exhibit similar defect severity levels. Incorporating this information will drive a departure from using binary wafer maps, which, by consequence, will require a new definition of what constitutes a cluster in our setting.

Appendix A
Additional Details About iWMM

Here we provide additional details about iWMM, which was initially proposed by [50]. iWMM comprises two building blocks: (1) a warping function to match the observed spatial locations of the AC-filtered results, denoted by $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_n]^T$, with a set of latent spatial coordinates in a latent space, denoted by $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]^T$, and (2) a clustering method which determines the clustering assignments in the latent space, denoted by $\mathbf{A} = [a_1, \ldots, a_n]^T$. While in theory, $\mathbf{z}_i$ can have a different dimensionality than $\mathbf{s}_i$, it suffices in our setting to assume that both $\mathbf{s}_i, \mathbf{z}_i \in \mathbb{R}^2$.

As a warping function, a Gaussian process latent variable model (GPLVM) [51] with squared exponential covariance, is used, and can be expressed as in Eq. (5). Then, the infinite Gaussian mixture model (iGMM) is used for spatial clustering, and is expressed as in Eq. (6). For iGMM, a Gaussian-Wishart prior is placed on its parameters $\boldsymbol{\mu}_k$ and $\mathbf{V}_k$, such that:

$$p\left(\boldsymbol{\mu}_k, \mathbf{V}_k\right) = \mathcal{N}\Big(\boldsymbol{\mu}_k | \mathbf{m}, (p\mathbf{V}_k)^{-1}\Big)\mathcal{W}\Big(\mathbf{V}_k | \mathbf{R}^{-1}, r\Big), \quad (13)$$

where $\mathcal{W}(\cdot)$ is the Wishart distribution. The parameters $\mathbf{m}, p$ are the mean and relative precision of $\boldsymbol{\mu}_k$, respectively, while

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.

$\mathbf{R}^{-1}$ and r are the scale matrix for $\mathbf{V}_k$, and its degree of freedom, respectively. One can then derive the probability distribution of $\mathbf{Z}$ given the clustering assignments $\mathbf{A}$ by integrating out $\boldsymbol{\mu}_k$ and $\mathbf{V}_k$, as in Eq. (14).

$$p(\mathbf{Z}|\mathbf{A}, \mathbf{R}, \mathbf{m}, r, p) = \prod_{k=1}^{\infty} \pi^{-n_k} \frac{p|\mathbf{R}|^{r/2}}{p_k|\mathbf{R}_k|^{r_k/2}} \times \prod_{j=1}^{2} \frac{\Gamma\left(\frac{r_k+1-j}{2}\right)}{\Gamma\left(\frac{r+1-j}{2}\right)}, \tag{14}$$

where n_k is the number of chips assigned to the kth sub-cluster, while p_k, r_k, and $\mathbf{S}_k$ are the posterior Gaussian-Wishart parameters of the kth component (or sub-cluster), such that $p_k = p + n_k$, $r_k = r + n_k$, and $\mathbf{S}_k = \mathbf{S} + \sum_{i:a_i=k} \mathbf{z}_i\mathbf{z}_i^T + p\mathbf{m}\mathbf{m}^T - p_k\mathbf{m}_k\mathbf{m}_k^T$, with $\mathbf{m}_k = \frac{p\mathbf{m}+\sum_{i:a_i=k}\mathbf{z}_i}{p+n_k}$.

A Dirichlet process prior with concentration parameter α is used for infinite mixture modeling in the latent space. Then, the probability distribution of $\mathbf{A}$ can be written as:

$$p(\mathbf{A}|\alpha) = \frac{\alpha^k \prod_{k=1}^{K}(n_k - 1)!}{\alpha(\alpha + 1)\cdots(\alpha + n - 1)}, \tag{15}$$

Collecting the above pieces, the joint distribution of $\mathbf{S}$, $\mathbf{Z}$, and $\mathbf{A}$ conditional on all parameters, can be written as:

$$p(\mathbf{S}, \mathbf{Z}, \mathbf{A}|\boldsymbol{\Theta}, \mathbf{R}, \mathbf{m}, r, p, \alpha) = p(\mathbf{S}|\mathbf{Z}, \boldsymbol{\Theta})p(\mathbf{Z}|\mathbf{A}, \mathbf{R}, \mathbf{m}, r, p) \times p(\mathbf{A}|\alpha), \tag{16}$$

which is merely the product of the terms determined by Eqs. (5), (14), and (15).

The authors in [50] provide a detailed procedure to fit the iWMM to a set of observed spatial locations $\mathbf{S}$, where the latent coordinates $\mathbf{Z}$, assignments $\mathbf{A}$, as well as remaining parameters are inferred through a Markov Chain Monte Carlo (MCMC)-based procedure. The procedure consists of two steps, which are repeatedly performed until convergence. The first step entails a Gibbs sampling scheme of the latent assignment of the ith chip, denoted by a_i, from the following probability distribution:

$$p\big(a_i = k|\mathbf{Z}, \mathbf{A}^{-i}, \mathbf{R}, \mathbf{m}, r, p, \alpha\big) \propto \begin{cases} n_k^{-i} p\Big(\mathbf{z}_i|\mathbf{Z}_k^{-i}, \mathbf{R}, \mathbf{m}, r, p\Big) & \text{assign to an existing sub-cluster} \\ \alpha p(\mathbf{z}_i|\mathbf{R}, \mathbf{m}, r, p) & \text{form a new sub-cluster,} \end{cases} \tag{17}$$

where $\mathbf{Z}_k^{-i}$ is the set of latent coordinates of the kth sub-cluster, excluding the ith chip. Similarly, $\mathbf{A}^{-i}$ is the set of assignments, excluding that of the ith chip, and n_k^{-i} is the number of chips assigned to the kth sub-cluster, excluding the ith chip. The probability distributions in the right hand-side of Eq. (17) can be analytically derived in closed-form as detailed in [50]. The second step entails sampling the latent coordinates $\mathbf{Z}$ from the probability distribution $p(\mathbf{Z}|\mathbf{A}, \mathbf{S}, \boldsymbol{\Theta}, \mathbf{R}, \mathbf{m}, r, p)$ using hybrid Monte Carlo. Combined, the two steps yield an estimate of the posterior distribution of the latent coordinates $\mathbf{Z}$ and the latent assignments $\mathbf{A}$.

Appendix B
Clustering Results for All Wafers

In Figure 10, we show the visual clustering results for all twelve wafers depicted in Figure 4.

Appendix C
Nomenclature

α	Concentration parameter for the Dirichlet process
β	The number of pairs pertaining to different sub-clusters in $\mathbf{A}$ and $\hat{\mathbf{A}}$
γ	The number of pairs pertaining to identical sub-clusters in $\mathbf{A}$ and $\hat{\mathbf{A}}$
$\hat{\mathbf{A}}$	The vector of predicted assignments of defective chips
$\hat{a}_i$	The predicted assignment of the ith defective chip
$\mathbf{A}$	The vector of true assignments of defective chips
$\mathbf{A}^{-i}$	The set of assignments, excluding the ith chip
$\mathbf{G}$	The center of all coordinates in $\mathbf{S}$
$\mathbf{G}^k$	The center of all coordinates in $\mathbf{S}^k$
$\mathbf{m}$	The mean parameter for $\boldsymbol{\mu}_k$
$\mathbf{R}^{-1}$	The scale matrix for $\mathbf{V}_k$
$\mathbf{S}$	The set of observed coordinates of the defective chips
$\mathbf{S}^k$	The chip coordinates of the kth sub-cluster
$\mathbf{s}_i$	The observed coordinates of the ith defective chip
$\mathbf{s}_i^k$	The ith chip's coordinates of the kth sub-cluster
$\mathbf{V}_k$	The precision matrix of the kth mixture component
$\mathbf{Z}$	The set of latent coordinates of the defective chips
$\mathbf{z}_i$	The latent coordinates of the ith defective chip
$\mathbf{Z}_k^{-i}$	The set of latent coordinates of the kth sub-cluster, excluding the ith chip
$\mathcal{W}(\cdot)$	The Wishart distribution
ϕ_k	The weight of the kth mixture component
$\boldsymbol{\mu}_k$	The mean of the kth mixture component
$\boldsymbol{\Sigma}$	The $n \times n$ covariance matrix whose ith and jth entry is the covariance between latent coordinates $\mathbf{z}_i$ and $\mathbf{z}_j$
$\boldsymbol{\Theta}$	Hyperparameters of the covariance function $C(\cdot)$
τ	The number of pairs pertaining to identical sub-clusters in $\mathbf{A}$ and different sub-clusters in $\hat{\mathbf{A}}$
ζ	The number of pairs pertaining to different sub-clusters in $\mathbf{A}$ and identical sub-clusters in $\hat{\mathbf{A}}$
$C(\cdot)$	A stationary parametric covariance function
d_i	The number of defects on the ith chip
E	The set of edges (pairs of adjacent chips)
$f_i(\cdot)$	Deviation cost function in the AC formulation
G	The graph representation of a wafer
$g_{ij}(\cdot)$	Separation cost function in the AC formulation
M	The pre-set threshold for the CPF method
n	The number of defective chips
n_k	The number of chips assigned to the kth sub-cluster
n_k^{-i}	The number of chips assigned to the kth sub-cluster, excluding the ith chip
p	The relative precision parameter for $\boldsymbol{\mu}_k$
r	The number of degrees of freedom for $\mathbf{V}_k$
u_{ij}	Separation cost of the ith and jth chips (for $[i, j] \in E$)

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.

V The set of nodes (chip indices)
w_i Deviation cost of the ith chip
x_i Cluster membership for the ith chip
z_{ij} Difference in cluster labels of the ith and jth chips
a_i The true assignment of the ith defective chip.

REFERENCES

[1] K. S. El-Kilany, "Reusable modelling and simulation of flexible manufacturing for next generation semiconductor manufacturing facilities," Ph.D. dissertation, Dept. School Mech. Manuf. Eng., Dublin City School, Dublin, OH, USA, 2003.

[2] J. Y. Hwang and W. Kuo, "Model-based clustering for integrated circuit yield enhancement," *Eur. J. Oper. Res.*, vol. 178, no. 1, pp. 143–153, 2007.

[3] J. Kim, Y. Lee, and H. Kim, "Detection and clustering of mixed-type defect patterns in wafer bin maps," *IISE Trans.*, vol. 50, no. 2, pp. 99–111, 2018.

[4] S. P. Cunningham and S. Mackinnon, "Statistical methods for visual defect metrology," *IEEE Trans. Semicond. Manuf.*, vol. 11, no. 1, pp. 48–53, Feb. 1998.

[5] S. Neo *et al.*, "Failure analysis methodology on donut pattern failure due to photovoltaic electrochemical effect," *Microelectron. Rel.*, vols. 76–77, pp. 255–260, Sep. 2017.

[6] G. Tello, O. Y. Al-Jarrah, P. D. Yoo, Y. Al-Hammadi, S. Muhaidat, and U. Lee, "Deep-structured machine learning model for the recognition of mixed-defect patterns in semiconductor fabrication processes," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 2, pp. 315–322, May 2018.

[7] K. Kyeong and H. Kim, "Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 3, pp. 395–402, Aug. 2018.

[8] Y. Kong and D. Ni, "Recognition and location of mixed-type patterns in wafer bin maps," in *Proc. IEEE Int. Conf. Smart Manuf. Ind. Logist. Eng. (SMILE)*, Hangzhou, China, 2019, pp. 4–8.

[9] H. Lee and H. Kim, "Semi-supervised multi-label learning for classification of wafer bin maps with mixed-type defect patterns," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 4, pp. 653–662, Nov. 2020.

[10] J. Wang, C. Xu, Z. Yang, J. Zhang, and X. Li, "Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 4, pp. 587–596, Nov. 2020.

[11] T. Yuan and W. Kuo, "A model-based clustering approach to the recognition of the spatial defect patterns produced during semiconductor fabrication," *IIE Trans.*, vol. 40, no. 2, pp. 93–101, 2007.

[12] T. Yuan and W. Kuo, "Spatial defect pattern recognition on semiconductor wafers using model-based clustering and Bayesian inference," *Eur. J. Oper. Res.*, vol. 190, no. 1, pp. 228–240, 2008.

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

[14] C. H. Jin, H. J. Na, M. Piao, G. Pok, and K. H. Ryu, "A novel DBSCAN-based defect pattern detection and classification framework for wafer bin map," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 3, pp. 286–292, Aug. 2019.

[15] C.-H. Wang, "Separation of composite defect patterns on wafer bin map using support vector clustering," *Expert Syst. Appl. Int. J.*, vol. 36, no. 2, pp. 2554–2561, 2009.

[16] L.-C. Chao and L.-I. Tong, "Wafer defect pattern recognition by multi-class support vector machines by using a novel defect cluster index," *Expert Syst. Appl.*, vol. 36, no. 6, pp. 10158–10167, 2009.

[17] W. Taam and M. Hamada, "Detecting spatial effects from factorial experiments: An application from integrated-circuit manufacturing," *Technometrics*, vol. 35, no. 2, pp. 149–160, 1993.

[18] Y.-S. Jeong, S.-J. Kim, and M. K. Jeong, "Automatic identification of defect patterns in semiconductor wafer maps using spatial correlogram and dynamic time warping," *IEEE Trans. Semicond. Manuf.*, vol. 21, no. 4, pp. 625–637, Nov. 2008.

[19] K. Taha, K. Salah, and P. D. Yoo, "Clustering the dominant defective patterns in semiconductor wafer maps," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 1, pp. 156–165, Feb. 2018.

[20] K. P. White, B. Kundu, and C. M. Mastrangelo, "Classification of defect clusters on semiconductor wafers via the hough transformation," *IEEE Trans. Semicond. Manuf.*, vol. 21, no. 2, pp. 272–278, May 2008.

[21] Y. Liu and S. Zhou, "Detecting point pattern of multiple line segments using hough transformation," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 1, pp. 13–24, Feb. 2015.

[22] B. Kim, Y.-S. Jeong, S. H. Tong, I.-K. Chang, and M.-K. Jeongyoung, "A regularized singular value decomposition-based approach for failure pattern classification on fail bit map in a dram wafer," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 1, pp. 41–49, Feb. 2015.

[23] R. Wang and N. Chen, "Wafer map defect pattern recognition using rotation-invariant features," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 596–604, Nov. 2019.

[24] M. Piao, C. H. Jin, J. Y. Lee, and J.-Y. Byun, "Decision tree ensemble-based wafer map failure pattern recognition based on radon transform-based features," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 2, pp. 250–257, May 2018.

[25] J. Yu and X. Lu, "Wafer map defect detection and recognition using joint local and nonlocal linear discriminant analysis," *IEEE Trans. Semicond. Manuf.*, vol. 29, no. 1, pp. 33–43, Feb. 2016.

[26] F. Adly, P. D. Yoo, S. Muhaidat, Y. Al-Hammadi, U. Lee, and M. Ismail, "Randomized general regression network for identification of defect patterns in semiconductor wafer maps," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 2, pp. 145–152, May 2015.

[27] M. Saqlain, B. Jargalsaikhan, and J. Y. Lee, "A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 2, pp. 171–182, May 2019.

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

[29] C.-T. Su, T. Yang, and C.-M. Ke, "A neural-network approach for semiconductor wafer post-sawing inspection," *IEEE Trans. Semicond. Manuf.*, vol. 15, no. 2, pp. 260–266, May 2002.

[30] F. Di Palma, G. De Nicolao, G. Miraglia, E. Pasquinetti, and F. Piccinini, "Unsupervised spatial pattern classification of electrical-wafer-sorting maps in semiconductor manufacturing," *Pattern Recognit. Lett.*, vol. 26, no. 12, pp. 1857–1865, 2005.

[31] C.-J. Huang, "Clustered defect detection of high quality chips using self-supervised multilayer perceptron," *Expert Syst. Appl.*, vol. 33, no. 4, pp. 996–1003, 2007.

[32] C.-W. Liu and C.-F. Chien, "An intelligent system for wafer bin map defect diagnosis: An empirical study for semiconductor manufacturing," *Eng. Appl. Artif. Intell.*, vol. 26, nos. 5–6, pp. 1479–1486, 2013.

[33] C.-F. Chien, S.-C. Hsu, and Y.-J. Chen, "A system for online detection and classification of wafer bin map defect patterns for manufacturing intelligence," *Int. J. Prod. Res.*, vol. 51, no. 8, pp. 2324–2338, 2013.

[34] H. Lee, Y. Kim, and C. O. Kim, "A deep learning model for robust wafer fault monitoring with sensor measurement noise," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 1, pp. 23–31, Feb. 2017.

[35] T. Nakazawa and D. V. Kulkarni, "Wafer map defect pattern classification and image retrieval using convolutional neural network," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 2, pp. 309–314, May 2018.

[36] N. Yu, Q. Xu, and H. Wang, "Wafer defect pattern recognition and analysis based on convolutional neural network," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 566–573, Nov. 2019.

[37] K. Imoto, T. Nakai, T. Ike, K. Haruki, and Y. Sato, "A CNN-based transfer learning method for defect classification in semiconductor manufacturing," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 455–459, Nov. 2019.

[38] Y. Kong and D. Ni, "A semi-supervised and incremental modeling framework for wafer map classification," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 1, pp. 62–71, Feb. 2020.

[39] J. Hwang and H. Kim, "Variational deep clustering of wafer map patterns," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 3, pp. 466–475, Aug. 2020.

[40] T.-H. Tsai and Y.-C. Lee, "A light-weight neural network for wafer map classification based on data augmentation," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 4, pp. 663–673, Nov. 2020.

[41] J. Jang, M. Seo, and C. O. Kim, "Support weighted ensemble model for open set recognition of wafer map defects," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 4, pp. 635–643, Nov. 2020.

[42] Y. Hyun and H. Kim, "Memory-augmented convolutional neural networks with triplet loss for imbalanced wafer defect pattern classification," *IEEE Trans. Semicond. Manuf.*, vol. 33, no. 4, pp. 622–634, Nov. 2020.

[43] D. S. Hochbaum and S. Liu, "Adjacency-clustering and its application for yield prediction in integrated circuit manufacturing," *Oper. Res.*, vol. 66, no. 6, pp. 1571–1585, 2018.

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.

[44] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton, FL, USA: Chapman and Hall, 2005.
[45] S. Geman and C. Graffigne, "Markov random field image models and their applications to computer vision," in *Proc. Int. Congr. Math.*, vol. 1. Berkeley, CA, USA, 1986, pp. 1496–1517.
[46] M. H. Hansen, V. N. Nair, and D. J. Friedman, "Monitoring wafer map data from integrated circuit fabrication processes for spatially clustered defects," *Technometrics*, vol. 39, no. 3, pp. 241–253, 1997.
[47] L. Bao, K. Wang, and R. Jin, "A hierarchical model for characterising spatial wafer variations," *Int. J. Prod. Res.*, vol. 52, no. 6, pp. 1827–1842, 2014.
[48] D. S. Hochbaum, "An efficient algorithm for image segmentation, Markov random fields and related problems," *J. ACM*, vol. 48, no. 4, pp. 686–701, Jul. 2001.
[49] R. K. Ahuja, D. S. Hochbaum, and J. B. Orlin, "Solving the convex cost integer dual network flow problem," *Manag. Sci.*, vol. 49, no. 7, pp. 950–964, 2003.
[50] T. Iwata, D. Duvenaud, and Z. Ghahramani, "Warped mixtures for nonparametric cluster shapes," 2012. [Online]. Available: arXiv:1206.1846
[51] N. D. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 2004, pp. 329–336.
[52] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
[53] T. Iwata and D. Duvenaud. (2016). *Code for Inferring Nonparametric Cluster Shapes Using Warped Mixtures*. [Online]. Available: https://github.com/duvenaud/warped-mixtures
[54] J.-S. R. Jang. (2014). *Wafer Map Dataset*. [Online]. Available: http://mirlab.org/dataSet/public/
[55] M.-J. Wu, J.-S. R. Jang, and J.-L. Chen, "Wafer map failure pattern recognition and similarity ranking for large-scale data sets," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 1, pp. 1–12, Feb. 2015.
[56] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognit.*, vol. 46, no. 1, pp. 243–256, 2013.
[57] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Stat. Theory Methods*, vol. 3, no. 1, pp. 1–27, 1974.
[58] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 28, no. 3, pp. 301–315, Jun. 1998.
[59] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.
[60] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.
[61] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *J. Mach. Learn. Res.*, vol. 11, pp. 2837–2854, Dec. 2010.
[62] M. J. Vavrek, "Fossil: Palaeoecological and palaeogeographical analysis tools," *Palaeontologia Electronica*, vol. 14, no. 1, pp. 1–16, 2011.
[63] B. Desgraupes, *Clustering Indices*, vol. 1, Univ. Paris Ouest-Lab Modal'X, Paris, France, 2013, p. 34.

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 07,2021 at 22:26:53 UTC from IEEE Xplore. Restrictions apply.