Spatial Pattern Recognition with Adjacency-Clustering: Improved Diagnostics for Semiconductor Wafer Bin Maps

Ahmed Aziz Ezzat\textsuperscript{1}, Sheng Liu\textsuperscript{2}, Dorit Hochbaum\textsuperscript{3}, Yu Ding\textsuperscript{4}

\textsuperscript{1}Department of Industrial & Systems Engineering, Rutgers University
\textsuperscript{2}Rotman School of Management, University of Toronto
\textsuperscript{3}Department of Industrial Engineering & Operations Research, University of California, Berkeley
\textsuperscript{4}Department of Industrial & Systems Engineering, Texas A&M University

In semiconductor manufacturing, statistical quality control hinges on an effective analysis 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 \textit{spatial pattern recognition}. Detecting defect patterns on a wafer can deliver key diagnostics about the root causes of defects and assist production engineers in mitigating future failures. Recently, there has been a growing interest in \textit{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 the filtered patterns into distinct defect types. Observing that spatial filtering is instrumental to high-quality 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 bin maps. Tested on real-world data and compared against a state-of-the-art approach, our proposed method achieves at least 49\% gain in terms of internal cluster validation quality (i.e., validation without external class labels), and about 6\% 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 defect pattern complexity, with larger gains achieved for more complex-shaped patterns. This superior performance is a testament to the method’s promising impact to semiconductor manufacturing, as well as other contexts where mixed-type spatial patterns are prevalent.

\textit{Key words}: Clustering, Pattern recognition, Semiconductor Manufacturing, Unsupervised learning

\section{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 \cite{El-Kilany2003}. 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 \textit{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.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{wafer_bin_maps.png}
\caption{Examples of wafer bin maps resulting from wafer probing tests. Panels (a-b) are \textit{single-type} patterns, i.e. one defect pattern per wafer, while Panel (c) shows a \textit{mixed-type} defect pattern. \textit{Random patterns} due to inherent process variation are represented by the scattered red squares on the wafer maps, which overlap with \textit{systematic patterns} (e.g. center, zone, donut) that are caused by assignable root causes.}
\end{figure}

A careful analysis of wafer bin maps is pivotal to quality control efforts in the semiconductor manufacturing industry. By examining 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. In a manufacturing process like wafer fabrication that entails hundreds of dependent and technologically complex steps, such analysis is of extreme importance to pinpoint possible root causes of failures. In fact, several spatial defect patterns in wafer bin maps can be directly
traced back 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 (Hwang and Kuo 2007, Kim et al. 2018), while a zone-shaped, edge-located defect pattern, as shown in Figure 1(b), can be traced back to uneven polishing or edge-die effects (Cunningham and Mackinnon 1998). A center-located, donut-shaped defect pattern, as shown in Figure 1(c), is routinely observed in wafer data due to possible tooling problems (Neo et al. 2017). These spatial patterns like center, zone, or donut, wherein defective chips “cluster” to form distinct shapes are referred to as **systematic patterns** since they are initiated by an assignable root cause. In contrast, randomly scattered defective chips in Figure 2(a)-(c) are referred to as **random patterns**, 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** patterns. The former refers to wafer maps that host only one defect pattern (e.g. center, zone, etc.), 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 wafer fabrication, mixed-type patterns are now increasingly observed in production data. Nevertheless, barring few recent efforts (Kim et al. 2018, Tello et al. 2018, Kong and Ni 2019), the problem of mixed-type SPR has received less attention than 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—See Figure 1, and (2) Spatial clustering, i.e., to group the filtered systematic patterns obtained from the first task into one or more sub-clusters pertaining to different types of defect patterns (e.g., center, zone, donut). A large body of literature has been devoted to improving the effectiveness of the second task, namely spatial clustering, among which those that are based on model-based clustering techniques (Yuan and Kuo 2007, 2008, Hwang and Kuo 2007, Yuan et al. 2011), kernel-based clustering (Wang 2009, Chao and Tong 2009), similarity-based metrics such as correlograms and nearest-neighbor measures of spatial dependence (Taam and Hamada 1993, Jeong et al. 2008), generalized linear mixed models (Wang et al. 2018), Hough transformations (White et al. 2008), and neural networks especially those based on adaptive resonance theory (Chen and Liu 2000, Su et al. 2002, Di Palma et al. 2005, Huang 2007, Liu and Chien 2013, Chien et al. 2013).
On the other hand, ad hoc heuristics are often used for the first task, namely spatial filtering, with an implicit assumption that the deficiencies of a poorly designed filtering step will be corrected in the second task. While this assumption may be acceptable for single-type SPR, we claim that the filtering step is of critical importance to mixed-type SPR. Motivated by a similar observation, Kim et al. (2018) propose an algorithm called connected path filtering (CPF) for the spatial filtering step to distinguish systematic and random defects in mixed-type wafer bin maps. This CPF algorithm is then coupled 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, $M$.

![Raw Wafer Map](Image)

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

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 certain globally pre-set threshold, $M$. In other words, if a chip is labeled as “functional,” while all of its neighbors are not, the CPF approach 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 which hosts a mixed-type pattern consisting of partial ring and donut defects. The results from the CPF method (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 (i.e., donut and partial ring), 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 2.

The second limitation of the CPF approach is its choice of the pre-set threshold $M$. Our findings, to be presented in Section 3, suggest that there does not appear to be a default value for $M$ that works universally well for all wafers and all combinations of defect types, making the choice of $M$ wafer-specific. Furthermore, as outlined by Kim et al. (2018), this choice is made via interaction with domain experts. This limitation severely hampers the applicability of the CPF approach in practice. Given the large production volumes typical of semiconductor production lines, the need for domain experts to constantly weigh in and update the value of the parameter $M$ can be extraordinarily inefficient and thus impractical.

Motivated by the need to address the above two limitations, we propose an innovative approach to extract mixed-type defect patterns in wafer bin maps. Our major finding is that a graph-theoretic approach which effectively leverages the local spatial dependence structure can have the potential to considerably improve the spatial filtering step, and ultimately, the overall quality of SPR. Specifically, we propose to use the adjacency-clustering (AC) model for spatial filtering, which was originally introduced by Hochbaum and Liu (2018) for yield prediction. Although technically similar, the main function of AC in our work is different from that in Hochbaum and Liu (2018): we focus on extracting systematic defect patterns (i.e., diagnostic analysis), rather than yield prediction (i.e., prognostic analysis). This distinction in motivation drives a fundamental departure in what defines a cluster from the earlier work by Hochbaum and Liu (2018): while Hochbaum and Liu (2018) define a cluster as a group of chips with homogeneous yield level, our approach, on the other hand, defines a cluster based on its chips’ membership in either the set of systematic or random defect pattern clusters. AC is, in fact, closely tied to Markov Random Field (MRF)
models in statistics and probability, which have been successfully applied in image segmentation and restoration [Rue and Held 2005, Geman and Graffigne 1986], spatial clustering detection [Hansen et al. 1997], and wafer thickness variation analysis [Bao et al. 2014].

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 proposed by Kim et al. (2018) by at least 49% in terms of internal cluster validation quality (i.e. validation without external information about class labels), and about 6% 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 Kim et al. (2018) 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. This superior performance is a testament to the method’s promising applicability to semiconductor manufacturing, as well as to other contexts in which mixed-type complex spatial patterns may be prevalent, such as energy networks and supply chain risk analysis.

Finally, we conclude this section by describing the organization of this paper. In Section 2, 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 3 presents our case study which details the analysis of nine real-world wafer maps exhibiting complex multi-type defect patterns. Section 4 concludes this paper and highlights future research directions.

2. 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 4 neighbors (rook-move neighborhood) or 8 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, the adjacency-clustering (AC) model 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 defect patterns (those caused by assignable root causes) from the random defect 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.

2.1. Spatial Filtering Stage: Adjacency-Clustering for Raw Wafer Data

We sketch here the adjacency-clustering model from Hochbaum and Liu (2018) and adapt it into our application. The adjacency-clustering 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 the assigned label \( x_i \) 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 \([i, j] \in E\). The adjacency-clustering model can be formulated as the following integer program:

\[
\begin{align*}
\min & \quad \sum_{i \in V} f_i(x_i, d_i) + \sum_{[i, j] \in E} g_{ij}(x_i - x_j) \\
\text{s.t.} & \quad x_i \in X \quad \forall \ i \in V,
\end{align*}
\] (AC)

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

The clustering result depends 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 (and 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 adjacency-clustering model has a statistics foundation from Markov random fields (MRF), and solving for $x_i$ is finding the maximum a posterior (MAP) estimate of the degradation model with an MRF prior (Geman and Graffigne 1986). Different forms of separation and deviation functions reflect different distributional assumptions of MRF.

The neighborhood system also plays an important role in the result from AC. 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 adjacency-clustering model reduces to the problem called minimum s-excess (Hochbaum 2001):

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

(AC-BIN)  

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

(2)  

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

(3)  

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

(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 minimum s-excess problem can be solved in polynomial time with a minimum-cut algorithm applied to an appropriately defined graph (Hochbaum 2001):
Proposition 1. The adjacency-clustering model with binary label values (AC-BIN) can be solved in polynomial time.

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 the real-time application of our framework.

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, Hochbaum (2001) 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 Hochbaum (2001) 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 (Ahuja et al. 2003).

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 unassignable 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). Our numerical analysis, to be presented in Section 3, suggests that there is a set of parameter values that yields consistently good results in our application of interest.

The resulting cluster pertaining to the systematic defects can contain more than one defect pattern (e.g., donut and partial ring—as in Figure 2). As such, we couple the AC method for spatial filtering in the first stage with a spatial clustering model in the second stage to group different patterns into sub-clusters. This spatial clustering model is introduced in subsection 2.2.
2.2. Spatial Clustering Stage: Infinite Warping Mixture Model for the AC-Filtered Results

Given the AC-filtered results, we apply the infinite warping mixture model (iWMM) \cite{iwata2012} to group the resulting systematic patterns into sub-clusters pertaining to distinct types of defect patterns (i.e. zone, donut, ring, etc—see Figure 1). 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 Iwata et al. \cite{iwata2012} and then adopted by Kim et al. \cite{kim2018} to spatially cluster the wafer maps that are filtered via their connected path filtering (CPF) approach. 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 Kim et al. \cite{kim2018}, we ensure that the improvements in clustering quality are mainly attributed to our proposed filtering method.

The benefit of applying iWMM in spatial clustering of defect patterns in wafer maps is two-fold. First, in iWMM, the number of sub-clusters corresponding to the number of defect types on a wafer (i.e., the number of sub-clusters) is estimated rather than specified \textit{a priori}—a common 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 distribution. 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 or (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. The clustering can then be performed in the latent space using a model-based (e.g., Gaussian) clustering technique. Figure 3 shows an example of how iWMM works within our SPR 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. \cite{iwata2012} propose to use a Gaussian process latent variable model (GPLVM) \cite{lawrence2006}...
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).

Using the notation from Section 3.1, we denote by \( S = [s_1, ..., s_n]^T \) the set of spatial locations associated with the defective chips in the AC-filtered results, i.e. for which \( x_i = 1 \), where \( n = \sum_i x_i \), and \( s_i \in \mathbb{R}^2 \). The set \( S \) in the observed space corresponds to a set of latent spatial coordinates in the latent space denoted by \( Z = [z_1, ..., z_n]^T \), where \( z_i \in \mathbb{R}^2 \). The ultimate goal of the iWMM is to find a vector of assignments in the latent space, denoted as \( A = [a_1, ..., a_n]^T \), where \( a_i \in \mathbb{Z}^+ \) denotes the membership of the \( i \)th chip to a particular sub-cluster.

The Gaussian process latent variable model (GPLVM) is used to map, or in other words, warp, the transformed spatial locations \( Z \), which are assumed to follow a Gaussian distribution, into the observable space where \( S \) can have a non-Gaussian distribution. For GPLVM, the conditional probability of \( S \) given \( Z \) is expressed as in Eq. (5) (Lawrence 2004).

\[
    p(S|Z, \Theta) = (2\pi)^{-n/2} |\Sigma|^{-1} \exp \left( -\frac{1}{2} \text{tr}(S^T \Sigma^{-1} S) \right),
\]

where \( \text{tr}(\cdot) \) is the trace function, \( |\cdot| \) is the determinant operation, and \( \Sigma \) is an \( n \times n \) covariance matrix whose \( i \)th and \( j \)th entry holds a covariance value, or in other words, a measure of similarity, between a pair of observations \( z_i \) and \( z_j \). A pillar question in GPLVMs, and GPs in general, is how to determine the entries of \( \Sigma \). Typically, a stationary parametric covariance function \( C(\cdot) \) is selected, which only depends on the lag between a pair of inputs through a set of hyperparameters \( \Theta \). A popular choice for \( C(\cdot) \) is the so-called squared exponential covariance (Williams and Rasmussen 2006), which is employed in this analysis.
Once the mapping (or warping) function is established, the infinite Gaussian mixture model (iGMM) is used for spatial clustering in the latent space. The iGMM assumes that the \( k \)th mixture component (or sub-cluster) in the latent space follows a Gaussian distribution with a mean and precision matrix, denoted by \( \mu_k \) and \( V_k \), respectively. In addition, each mixture component is associated with a mixture weight denoted by \( \phi_k \). The mathematical expression for iGMM is presented in Eq. (6).

\[
p(z|\phi_k, \mu_k, V_k) = \sum_{k=1}^{\infty} \phi_k N(z|\mu_k, V_k^{-1})
\] (6)

Iwata et al. (2012) provide a detailed procedure to fit the iWMM to a set of observed spatial locations \( S \), where the latent coordinates \( Z \), assignments \( A \), 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 (Iwata and Duvenaud 2016) and we use them for our numerical analysis in Section 3.

3. Application to Real-World Wafer Map Data

In this section, we evaluate the performance of our proposed two-stage SPR approach on a set of real-world wafer bin maps. We demonstrate its superior performance using a collection of widely recognized clustering quality metrics. We then proceed to infer key insights about its performance relative to a state-of-the-art spatial filtering method.

3.1. Data Description

We extracted nine wafer maps from a public dataset that is widely cited in the semiconductor manufacturing community (Jang 2014, Wu et al. 2014). While the original dataset contains a large number of wafer bin maps, we have selected nine wafers so as to (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 Kim et al. (2018) (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 nine chosen wafer maps, where the red- and green-colored squares represent the locations of defective and functional chips, respectively. Panels (a), (f) and (g) show wafer maps with center and partial ring defects. Panel (b) shows a wafer map with center and zone defects. Panels (c), (d), and (e) show wafer maps with two zone defects, while Panels (h) and (i) show wafer maps with donut and partial ring defects.
Figure 4  Nine wafer maps with mixed-type defects. Panels (a), (f) and (g) show wafer maps with center and partial ring defects. Panel (b) shows a wafer map with center and zone defects. Panels (c), (d), and (e) show wafer maps with two zone defects, while Panels (h) and (i) show wafer maps with donut and partial ring defects.

3.2. 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 benchmarking alternative in the comparison is the state-of-the-art approach in the semiconductor manufacturing literature proposed by Kim et al. (2018), 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 the downstream spatial clustering 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 Hochbaum and Liu (2018). As discussed in subsection 2.1, having a too large or too small value of $u$ is not ideal for the spatial filtering purpose. We observe that for the problem of SPR in wafer bin maps, the choice of $u = 0.5$ achieves a sensible trade-off between the deviation and separation costs, and yields consistently superior performance. Hence we only present the AC-iWMM results for $u = 0.5$. 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 pattern types. To showcase this point, we present the results of the CPF method at two different values of $M$ (findings to be discussed in the sequel). We compare AC-iWMM against CPF-iWMM on the nine wafer maps of Figure 4. We have implemented the CPF algorithm following the description of the method by Kim et al. (2018), while for iWMM, we have adapted the code made publicly available by Iwata and Duvenaud (2016). Figure 5 shows the results of the AC-iWMM approach for four wafer maps, starting from the raw wafer map, to AC-filtering result, to the iWMM clustering result in both the original and latent spaces. The visual results for all wafers are shown in Figure 11 in the Appendix.

3.2.1. 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. The middle and right panels show the results from the AC-iWMM and the CPF-iWMM, respectively. 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 on its own, as it does not utilize the local neighborhood information. We stress that iWMM was run at the same initial values for the parameters and the same number of iterations in both AC-iWMM and CFP-iWMM, so the difference between the two sets of results is solely attributed to the different spatial filtering approach (adjacency-clustering versus connected path filtering)
Figure 5  Results from four out of nine selected wafers, starting from the raw wafer maps (first column), to the AC filtering results using $u = 0.5$, and then to the iWMM clustering as applied to the AC-filtered data, in both the original and latent spaces (third and fourth columns, respectively).

Another illustrative example is shown in Figure 7 in which the wafer map exhibits two overlaid zone defects. Similar to the example in Figure 6, 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 the AC-iWMM approach that yields a clear distinction between the two sub-clusters corresponding to the two zone defect patterns. We note that this problem cannot be alleviated by just adjusting the value of the pre-set threshold $M$ because the set of chips that are mistakenly flagged by CPF-iWMM as a separate sub-cluster is 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
Visual comparison of CPF-iWMM (Panel b) and AC-iWMM (Panel c) on a wafer map that exhibits donut and partial ring defects. In contrast to AC, CPF fails to clearly separate the two sets of defective chips, causing iWMM to mistakenly flag a new separate sub-cluster. In Section 3.2.2, we further demonstrate that our approach outperforms the CPF-iWMM at various values of the threshold $M$.

Visual comparison of CPF-iWMM (Panel b) and AC-iWMM (Panel c) on a wafer map that exhibits two zone defects. In contrast to AC, CPF fails to clearly separate the two sets of defective chips, causing iWMM to mistakenly flag a new separate sub-cluster.

3.2.2. Quantitative Comparisons: The clustering results obtained by both approaches are then evaluated on two sets of performance metrics that are known in the pattern recognition literature as internal and external indices [Arbelaitz et al. 2013]. Assuming $\hat{A} = [\hat{a}_1, ..., \hat{a}_n]^T$ and $A = [a_1, ..., a_n]^T$ are the sets of predicted and true cluster assignments, respectively, internal indices assess the quality of the pattern recognition results when the underlying ground truth is not
available, that is, without access to the set $A$. External indices, on the other hand, make use of external information, namely the set $A$, to validate the estimated pattern recognition results.

Let us denote by $G$ the center of all coordinates in $S$. Similarly, $G_k$ denotes the center of the coordinates in $S_k = \{ s_i^k \}_{i: \hat{a}_i = k}$, i.e., the coordinates of the observations assigned to the $k$th sub-cluster, that is, 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) (Caliński and Harabasz 1974). By definition, a higher value for the CH index indicates a better performance.

$$CH(S, S^1, ..., S^K) = \frac{n - \hat{K}}{K - 1} \frac{\sum_{k=1}^{\hat{K}} n_k ||G^k - G||^2}{\sum_{k=1}^{\hat{K}} \sum_{i: \hat{a}_i = k} ||s_i^k - G_k||^2},$$

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) (Bezdek and Pal 1998). A higher value for GDI indicates a better performance.

$$GDI(S^1, ..., S^K) = \min_{k \neq k'} \frac{1}{n_k + n_{k'}} \left( \sum_{i: \hat{a}_i = k} ||s_i^k - G_k|| + \sum_{j: \hat{a}_j = k'} ||s_j^{k'} - G_k'|| \right) \max_k \max_{i \neq j: \hat{a}_i = \hat{a}_j} \frac{||s_i^k - s_j^{k'}||}{||s_i^k - G_k||}.$$

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 validity of the competing pattern recognition approaches. This is the approach adopted for benchmarking and evaluation in Kim et al. (2018). Since our dataset does not have the “ground truth,” or in other words the set $A$, we revert to the pattern recognition literature to reconstruct the ground truth from the raw data by applying a pattern reconstruction technique which iterates over every pixel of an image and updates its value using a weighted sum of its surrounding pixels to generate an output image (Gonzalez and Woods 2002). For our application, we used a $3 \times 3$ neighborhood system with $\frac{4}{9}$ weight (we observed the weight selection has minimal impacts on the final results).

The three external indices used in Kim et al. (2018) 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 $i$th sub-cluster of $A$ and the $j$th sub-cluster of $\hat{A}$. Now, let us define $\gamma$ as the number of pairs pertaining to the same sub-cluster in $A$ and to the same sub-cluster in $\hat{A}$, while $\beta$, on the other hand, denotes the number of pairs pertaining to different sub-clusters in $A$ and different sub-clusters in $\hat{A}$. With the above notations, RI can be defined as (Rand 1971),

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

(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:

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

(10)

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

The third external metric is NMI (Vinh et al. 2010), which is an information-theoretic metric that measures the amount of information that $A$ and $\hat{A}$ share, and is expressed as in Eq. (12). NMI ranges between 0 and 1, with higher values indicating better performance.

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

(11)

such that

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

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

(12)

Tables 1 and 2 summarize the comparison results, in terms of internal and external metrics, respectively, for the nine wafer maps of Figure 4. We have included results of the CPF approach at two different values of the threshold, namely $M = 5$ and $M = 10$. 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 (Vavrek 2011), while NMI is computed by calling the library NMI. All internal indices are computed by using functionalities in the library clusterCrit (Desgraupes 2013).
As shown in Tables 1 and 2, 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 Kim et al. (2018)), which can considerably compromise the usage of this approach in practice. As opposed to CPF, the choice of $u = 0.5$ for AC is shown to provide consistently satisfactory performance across all the wafers.

|   | Internal Indices |   | External Indices |   |
|---|------------------|---|------------------|---|
|   | CH   | GDI  | RI   | ARI  | NMI  |
| Wafer # | $M = 5$ | $M = 10$ | $M = 5$ | $M = 10$ | $M = 5$ | $M = 10$ | $M = 5$ | $M = 10$ | $M = 5$ | $M = 10$ | $M = 5$ | $M = 10$ |
| 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% | -0.36% | -0.36% | 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% |   |
| Avg. | 68.8% | 49.9% | 75.3% | 67.2% | 2.67% | 2.67% | 6.32% | 6.37% | 6.14% | 5.67% |   |

Table 3 Improvement (in percentage) of AC-iWMM over CPF-iWMM with $M = 5, 10$, for all metrics (internal and external) across the 9 wafers. Red-coloured values denote instances where percentage improvement was negative.

Table 3 presents the percentage improvements of AC-iWMM relative to CPF-iWMM at $M = 5, 10$, for all metrics. On average (last row of Table 3), AC-iWMM achieves an average improvement of up to 75% 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. Figure 8 shows the boxplots of all five metrics (internal and external) across the nine wafers, which further confirm the considerable improvement brought by AC-iWMM.

Figure 8  Top row: Boxplots across all wafers for internal metrics, namely: (a) CH (higher the better) and (b) GDI (higher the better). Bottom row: External metrics, namely: (c) RI (higher the better), (d) ARI (higher the better), and (e) NMI (higher the better).

Another interesting observation to highlight is the magnitude of improvement realized by AC-iWMM over CPF-iWMM as a function of the complexity of the defect patterns in the wafer map. Specifically, we note that improvements from our proposed approach are more pronounced for more complex-shaped defect patterns, and are diminishing as the defect pattern shapes become relatively simpler. For instance, considerable improvements (and in fact, maximal for some metrics)
in Tables 1 and 2 come from wafer map #9, hosting donut and partial ring defect patterns. The results for that particular wafer map is shown in Figure 9. Understandably, donut and partial ring defect patterns are, by design, intricate shapes, making the distinction between the 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.

![Figure 9](image)

**Figure 9** Visual comparison of AC-iWMM and CPF-iWMM on wafer map #9, which exhibits donut and partial ring defects. We note that CPF fails to clearly separate the two sets of defective chips, causing iWMM to mistakenly flag a new separate sub-cluster. In this wafer, AC-iWMM achieve considerable improvements over CPF-iWMM owing to its ability to better filter complex-shaped defect patterns.

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 were able to produce satisfactory performance, with almost negligible visual differences, as shown in Figure 10.

The observations from Figures 9 (for wafer map #9) and 10 (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 3 suggests the same observation for wafer maps #6 (visual result shown in Figure 11 in the Appendix) and #8 (visual result shown in Figure 6), which have complex-shaped patterns, and hence, the benefit
Visual comparison of AC-iWMM and CPF-iWMM on wafer map #4 with two zone defects. We note that both approaches render very similar results, visually, and quantitatively. The marginal difference in performance is due to the simplicity of the defect patterns—two zone defects with sparse random defects in the background, which are effectively filtered by both AC-iWMM and CPF-iWMM.

from AC-iWMM appears to be more pronounced. As the wafer fabrication process grows in scale and sophistication, owing to technology upgrade, 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. Therefore, we anticipate that our proposed approach will generate more profound impacts.

4. Conclusions

In this paper we have proposed an innovative pattern recognition framework (AC-iWMM) for spatial data with mixed-type patterns. This framework integrates the adjacency-clustering model for spatial filtering with an advanced mixture model (iWMM) for spatial clustering. The adjacency-clustering model, enabled by utilizing the local neighborhood information, is able to distinguish the systematic patterns from random noises. The adjacency-clustering model has a desirable combinatorial structure and can be solved in polynomial time by a minimum-cut algorithm. Because of this effective spatial filtering, iWMM, which subsequently acts on the filtered data, can properly cluster the systematic patterns into different types. We validate the superior performance of AC-iWMM in detecting mixed-type defect patterns in wafer bin maps—a problem of vital importance to ensuring quality control in the semiconductor manufacturing industry. Based on both visual and quantitative comparisons in clustering quality, AC-iWMM outperforms the state-of-the-art method in the literature, especially for complex-shaped, mixed-type patterns.
With the development of the Internet of Things (IoT), spatial data streams are continuously generated through sensors deployed in different locations (Feng and Shanthikumar 2018). Our proposed AC-iWMM framework can be applied to other types of spatial data, where multiple spatial patterns co-exist. For example, in public health sciences, AC-iWMM can be employed to analyze spatial epidemiological data by automatically identifying different transmission patterns that require immediate interventions. Also, in the context of spatial demand monitoring (e.g., delivery requests), our approach can be used to locate, in real-time, abrupt demand patterns due to special events like concerts or sports. Similarly, in energy network management, our approach can be adapted to monitor the energy consumption levels in different communities. We plan to investigate more applications of AC-iWMM in the future.

Acknowledgments

The research of Dorit S. Hochbaum is supported in part by NSF award No. CMMI-1760102.

References

Ahuja, Ravindra K, Dorit S Hochbaum, James B Orlin. 2003. Solving the convex cost integer dual network flow problem. Management Science 49(7) 950–964.

Arbelaitz, Olatz, Ibai Gurrutxaga, Javier Muguerza, Jesús M Pérez, Iñigo Perona. 2013. An extensive comparative study of cluster validity indices. Pattern Recognition 46(1) 243–256.

Bao, Lulu, Kaibo Wang, Ran Jin. 2014. A hierarchical model for characterising spatial wafer variations. International Journal of Production Research 52(6) 1827–1842.

Bezdek, James C, Nikhil R Pal. 1998. Some new indexes of cluster validity. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 28(3) 301–315.

Caliński, Tadeusz, Jerzy Harabasz. 1974. A dendrite method for cluster analysis. Communications in Statistics-Theory and Methods 3(1) 1–27.

Chao, Li-Chang, Lee-Ing Tong. 2009. Wafer defect pattern recognition by multi-class support vector machines by using a novel defect cluster index. Expert Systems with Applications 36(6) 10158–10167.

Chen, Fei-Long, Shu-Fan Lin. 2000. A neural-network approach to recognize defect spatial pattern in semiconductor fabrication. IEEE Transactions on Semiconductor Manufacturing 13(3) 366–373.
Chien, Chen-Fu, Shao-Chung Hsu, Ying-Jen Chen. 2013. A system for online detection and classification of wafer bin map defect patterns for manufacturing intelligence. *International Journal of Production Research* **51**(8) 2324–2338.

Cunningham, Sean P, Scott Mackinnon. 1998. Statistical methods for visual defect metrology. *IEEE Transactions on Semiconductor Manufacturing* **11**(1) 48–53.

Desgraupes, Bernard. 2013. Clustering indices. *University of Paris Ouest-Lab ModalX* **1** 34.

Di Palma, Federico, Giuseppe De Nicolao, Guido Miraglia, Egidio Pasquinetti, Francesco Piccinini. 2005. Unsupervised spatial pattern classification of electrical-wafer-sorting maps in semiconductor manufacturing. *Pattern Recognition Letters* **26**(12) 1857–1865.

El-Kilany, Khaled S. 2003. Reusable modelling and simulation of flexible manufacturing for next generation semiconductor manufacturing facilities. Ph.D. thesis, Dublin City School. School of Mechanical and Manufacturing Engineering.

Feng, Qi, J George Shanthikumar. 2018. How research in production and operations management may evolve in the era of big data. *Production and Operations Management* **27**(9) 1670–1684.

Geman, Stuart, Christine Graffigne. 1986. Markov random field image models and their applications to computer vision. *Proceedings of the International Congress of Mathematicians*, vol. 1. Berkeley, CA, 1496–1517.

Gonzalez, Rafael C, Richard E Woods. 2002. *Digital Image Processing*. Prentice Hall.

Hansen, Mark H, Vijayan N Nair, David J Friedman. 1997. Monitoring wafer map data from integrated circuit fabrication processes for spatially clustered defects. *Technometrics* **39**(3) 241–253.

Hochbaum, Dorit S. 2001. An efficient algorithm for image segmentation, markov random fields and related problems. *Journal of the ACM (JACM)* **48**(4) 686–701.

Hochbaum, Dorit S, Sheng Liu. 2018. Adjacency-clustering and its application for yield prediction in integrated circuit manufacturing. *Operations Research* **66**(6) 1571–1585.

Huang, Chenn-Jung. 2007. Clustered defect detection of high quality chips using self-supervised multilayer perceptron. *Expert Systems with Applications* **33**(4) 996–1003.
Hwang, Jung Yoon, Way Kuo. 2007. Model-based clustering for integrated circuit yield enhancement. *European Journal of Operational Research* **178**(1) 143–153.

Iwata, Tomoharu, David Duvenaud. 2016. Code for inferring nonparametric cluster shapes using warped mixtures. [https://github.com/duvenaud/warped-mixtures](https://github.com/duvenaud/warped-mixtures).

Iwata, Tomoharu, David Duvenaud, Zoubin Ghahramani. 2012. Warped mixtures for nonparametric cluster shapes. *arXiv preprint arXiv:1206.1846*.

Jang, Jyh-Shing R. 2014. Wafer map dataset. [http://mirlab.org/dataSet/public/](http://mirlab.org/dataSet/public/)

Jeong, Young-Seon, Seong-Jun Kim, Myong K Jeong. 2008. Automatic identification of defect patterns in semiconductor wafer maps using spatial correlogram and dynamic time warping. *IEEE Transactions on Semiconductor Manufacturing* **21**(4) 625–637.

Kim, Jinho, Youngmin Lee, Heeyoung Kim. 2018. Detection and clustering of mixed-type defect patterns in wafer bin maps. *IISE Transactions* **50**(2) 99–111.

Kong, Yuting, Dong Ni. 2019. Recognition and location of mixed-type patterns in wafer bin maps. *2019 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE)*. IEEE, 4–8.

Lawrence, Neil D. 2004. Gaussian process latent variable models for visualisation of high dimensional data. *Advances in Neural Information Processing Systems*. 329–336.

Liu, Chiao-Wen, Chen-Fu Chien. 2013. An intelligent system for wafer bin map defect diagnosis: An empirical study for semiconductor manufacturing. *Engineering Applications of Artificial Intelligence* **26**(5-6) 1479–1486.

Neo, SP, ACT Quah, GB Ang, D Nagalingam, HH Ma, SL Ting, CW Soo, CQ Chen, ZH Mai, JC Lam. 2017. Failure analysis methodology on donut pattern failure due to photovoltaic electrochemical effect. *Microelectronics Reliability* **76** 255–260.

Rand, William M. 1971. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* **66**(336) 846–850.

Rue, Havard, Leonhard Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC.
Su, Chao-Ton, Taho Yang, Chir-Mour Ke. 2002. A neural-network approach for semiconductor wafer post-sawing inspection. *IEEE Transactions on Semiconductor Manufacturing* **15**(2) 260–266.

Taam, Winson, Michael Hamada. 1993. Detecting spatial effects from factorial experiments: An application from integrated-circuit manufacturing. *Technometrics* **35**(2) 149–160.

Tello, Ghalia, Omar Y Al-Jarrah, Paul D Yoo, Yousof Al-Hammadi, Sami Muhaidat, Uihyoung Lee. 2018. Deep-structured machine learning model for the recognition of mixed-defect patterns in semiconductor fabrication processes. *IEEE Transactions on Semiconductor Manufacturing* **31**(2) 315–322.

Vavrek, Matthew J. 2011. Fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica* **14**(1) 1–16.

Vinh, Nguyen Xuan, Julien Epps, James Bailey. 2010. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research* **11** 2837–2854.

Wang, Chih-Hsuan. 2009. Separation of composite defect patterns on wafer bin map using support vector clustering. *Expert Systems with Applications* **36**(2) 2554–2561.

Wang, Hao, Bo Li, Seung Hoon Tong, In-Kap Chang, Kaibo Wang. 2018. A discrete spatial model for wafer yield prediction. *Quality Engineering* **30**(2) 169–182.

White, K Preston, Bijoy Kundu, Christina M Mastrangelo. 2008. Classification of defect clusters on semiconductor wafers via the hough transformation. *IEEE Transactions on Semiconductor Manufacturing* **21**(2) 272–278.

Williams, Christopher K.I., Carl Edward Rasmussen. 2006. *Gaussian Processes for Machine Learning*. MIT Press Cambridge, MA.

Wu, Ming-Ju, Jyh-Shing R Jang, Jui-Long Chen. 2014. Wafer map failure pattern recognition and similarity ranking for large-scale data sets. *IEEE Transactions on Semiconductor Manufacturing* **28**(1) 1–12.

Yuan, Tao, Way Kuo. 2007. A model-based clustering approach to the recognition of the spatial defect patterns produced during semiconductor fabrication. *IIE Transactions* **40**(2) 93–101.

Yuan, Tao, Way Kuo. 2008. Spatial defect pattern recognition on semiconductor wafers using model-based clustering and Bayesian inference. *European Journal of Operational Research* **190**(1) 228–240.
Yuan, Tao, Way Kuo, Suk Joo Bae. 2011. Detection of spatial defect patterns generated in semiconductor fabrication processes. *IEEE Transactions on Semiconductor Manufacturing* 24(3) 392–403.

**Appendix**

A1. **Additional details about the infinite warping mixture model (iWMM)**

Here we provide additional details about iWMM, which was initially proposed by [Iwata et al. 2012](#) for nonparametric clustering. In this paper, we use iWMM to group the resulting systematic defects filtered by the AC model into sub-clusters pertaining to distinct types of defect patterns. This is the same spatial clustering model used by [Kim et al. 2018](#). iWMM comprises two building blocks: (1) a warping function to match the observed spatial locations of the AC-filtered results, denoted by \( S = [s_1, \ldots, s_n]^T \), with a set of latent spatial coordinates in a latent space, denoted by \( Z = [z_1, \ldots, z_n]^T \), and (2) a model-based clustering method which determines the clustering assignments in the latent space, denoted by \( A = [a_1, \ldots, a_n]^T \). While in theory, \( z_i \) can have a different dimensionality than \( s_i \), it suffices in our setting to assume that both \( s_i, z_i \in \mathbb{R}^2 \).

As a warping function, a Gaussian process latent variable model (GPLVM) ([Lawrence 2004](#)), 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 the iGMM, a Gaussian-Wishart prior is placed on its parameters \( \mu_k \) and \( V_k \), such that:

\[
p(\mu_k, V_k) = N(\mu_k | m, (pV_k)^{-1}) \mathcal{W}(V_k | R^{-1}, r),
\]

(13)

where \( \mathcal{W}(\cdot) \) is the Wishart distribution. The parameters \( m, p \) are the mean and relative precision of \( \mu_k \), respectively, while \( R^{-1} \) and \( r \) are the scale matrix for \( V_k \), and its degree of freedom, respectively. One can then derive the probability distribution of \( Z \) given the clustering assignments \( A \) by integrating out \( \mu_k \) and \( V_k \), as in Eq. (14).

\[
p(Z | A, R, m, r, p) = \prod_{k=1}^{\infty} \pi^{-n_k} \frac{p|R|^r}{p_k|R_k|^r} \times \prod_{j=1}^{2} \frac{\Gamma(n_k + 1 - j)}{\Gamma(\frac{r + 1 - j}{2})}.
\]

(14)

where \( n_k \) is the number of chips assigned to the \( k \)th sub-cluster, while \( p_k, r_k \), and \( S_k \) are the posterior Gaussian-Wishart parameters of the \( k \)th component (or sub-cluster), such that \( p_k = p + n_k \), \( r_k = r + n_k \), and \( S_k = S + \sum_{i: a_i = k} z_i z_i^T + pmm^T - p_k m_k m_k^T \), with \( m_k = \frac{p_m + \sum_{i: a_i = k} z_i}{p + n_k} \).
A Dirichlet process prior with concentration parameter \( \alpha \) is used for infinite mixture modeling in the latent space. Then, the probability distribution of \( A \) can be written as:

\[
p(A|\alpha) = \frac{\alpha^k \prod_{k=1}^{K} (n_k - 1)!}{\alpha(\alpha + 1)\ldots(\alpha + n - 1)},
\]

(15)

Collecting the above pieces, the joint distribution of \( S, Z, \) and \( A \) conditional on all parameters, can be written as in Eq. (16):

\[
p(S, Z, A|\Theta, R, m, r, p, \alpha) = p(S|Z, \Theta)p(Z|A, R, m, r, p)p(A|\alpha),
\]

(16)

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

Iwata et al. (2012) provide a detailed procedure to fit the iWMM to a set of observed spatial locations \( S \), where the latent coordinates \( Z \), assignments \( A \), 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 \( i \)th chip, denoted by \( a_i \), from the following probability distribution:

\[
p(a_i = k|Z, A^{-i}, R, m, r, p, \alpha) \propto \begin{cases} 
n_k^i p(z_i|Z_k^{-i}, R, m, r, p) & \text{assign to an existing sub-cluster} \\
\alpha p(z_i|R, m, r, p) & \text{form a new sub-cluster},
\end{cases}
\]

(17)

where \( Z_k^{-i} \) is the set of latent coordinates assigned to the \( k \)th sub-cluster, excluding the \( i \)th chip. Similarly, the set \( A^{-i} \) is the set of chip assignments, excluding that of the \( i \)th chip, and \( n_k^i \) is the number of chips assigned to the \( k \)th sub-cluster, excluding the \( i \)th chip. The probability distributions in the right hand-side of Eq. (17) can be analytically derived in closed-form as detailed by Iwata et al. (2012). The second step entails sampling the latent coordinates \( Z \) from the probability distribution \( p(Z|A, S, \Theta, R, m, r, p) \) using hybrid Monte Carlo. Combined, the two steps yield an estimate of the posterior distribution of the latent coordinates \( Z \) and the latent assignments \( A \).

A2. Clustering results for all wafers

Here, we present the full visual results for the wafer clustering of the nine wafers shown in Figure 4. The results shown earlier in Figure 5 represent a subset of the full results shown in Figure 11. Across all nine wafers, AC-iWMM was able to clearly provide satisfactory performance in filtering the raw wafer bin maps, and ultimately distinguishing different defect patterns.
| Wafer # | Raw Wafer Map | AC ($u = 0.5$) | AC-iWMM (observed space) | AC-iWMM (latent space) |
|---------|---------------|----------------|--------------------------|------------------------|
| 1       | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image4.png) |
| 2       | ![Image](image5.png) | ![Image](image6.png) | ![Image](image7.png) | ![Image](image8.png) |
| 3       | ![Image](image9.png) | ![Image](image10.png) | ![Image](image11.png) | ![Image](image12.png) |
| 4       | ![Image](image13.png) | ![Image](image14.png) | ![Image](image15.png) | ![Image](image16.png) |
| 5       | ![Image](image17.png) | ![Image](image18.png) | ![Image](image19.png) | ![Image](image20.png) |
| 6       | ![Image](image21.png) | ![Image](image22.png) | ![Image](image23.png) | ![Image](image24.png) |
| 7       | ![Image](image25.png) | ![Image](image26.png) | ![Image](image27.png) | ![Image](image28.png) |
| 8       | ![Image](image29.png) | ![Image](image30.png) | ![Image](image31.png) | ![Image](image32.png) |
| 9       | ![Image](image33.png) | ![Image](image34.png) | ![Image](image35.png) | ![Image](image36.png) |

**Figure 11** Results from all nine wafers, starting from wafer id (first column), raw wafer maps (second column), to the AC filtering results (third column), and then to the iWMM clustering as applied to the AC-filtered data, in both the original and latent spaces (fourth and fifth columns, respectively).