ABSTRACT

An Information-Theoretic Measure for Pattern Similarity in Analog Wafermaps

Bernhard C. Geiger[1, *], Stefan Schrunner[2, *], Roman Kern[1]

[1] **Know-Center GmbH, Graz, AT**
[2] **KAI Kompetenzzentrum Automobil- und Industrieelektronik GmbH, Villach, AT**
[*] equal contribution

Presenter: Bernhard C. Geiger
E-mail: bgeiger@know-center.at

Presentation (oral and/or poster**): oral**

TOPIC:
Please select one or more fitting topic(s) for your contribution from the list. (1=first, 2=second etc. choice)

Process Level APC	**Fab Level APC**	**Manufacturing Effectiveness and Productivity**
□ Plasma etch, CVD and ALD	□ Fab level process control methods	□ Unit process & equipment productivity
□ Sputtering, P3I, and e--beam	□ Virtual metrology	1 Factory productivity and automation
□ Lithography	□ Yield management	□ Factory modeling, simulation and optimization
□ Thermal, wet processing & CMP	2 Factory data analysis	□ Cost optimization and end-of-life equipment issues
□ Backend	□ IT infrastructure	□ Environment and Green Manufacturing
□ Metrology and R2R		
□ APC for legacy tools		

ABSTRACT

Motivation

The development of automated tools for data analysis is a growing importance in semiconductor manufacturing. For this purpose, especially analog wafer test data carry a high potential as they allow the expert to draw conclusions on both product performance and production process deviations even before individual devices violate specifications. More specifically, it is assumed that process deviations are identifiable by spatial patterns (e.g., lines, rings, etc.) on the analog wafermaps generated from these wafer test data.

The target of this work is to support automated systems that cluster or classify wafermaps depending on the depicted process patterns. For this purpose, we propose a wafermap similarity measure that is based on empirical distributions extracted from the wafermaps.

Description

Suppose that each wafermap consists of n devices which are characterized by numerical features taking values in a finite set. We can thus represent each wafermap by a normalized histogram of these numerical features (cf. Santos et al., 2019, Sec. III.A), which can be interpreted as an empirical discrete probability distribution. Given that a pattern is sufficiently captured by these features, two wafermaps W_1 and W_2 are assumed to be similar to the extent to which their empirical probability distributions p_1 and p_2 are computed to be similar. We measure the similarity between discrete probability distributions (thus between wafermaps) using the **Jensen-Shannon divergence (JSD)** (Lin, 1991)

$$JSD(W_1, W_2) := H\left(\frac{p_1+p_2}{2}\right) - 0.5H(p_1) - 0.5H(p_2),$$

where $H(p)$ is the entropy of the discrete distribution p (Cover & Thomas, 1991). The JSD is non-negative, where small values indicate that the two wafermaps are similar, and has been used for image processing applications in the past, e.g., (Gómez-Lopera et al, 2000).

To meaningfully employ JSD as a measure of pattern similarity, one has to ensure that the empirical probability distribution of the device features in some sense represents the pattern. We utilize the feature extraction scheme proposed in (Santos et al., 2019) for analog wafermaps: In the first step, the wafer W is segmented into regions of interest R. Each device within these regions is represented by a feature obtained by comparing its measurement value to those of its neighbors. Finally, the empirical probability distribution of the wafer W is obtained by normalizing the histogram of these feature values by the number of devices $|R|$ on the region of interest.

We make two simplifications: First, we assume that outside the regions of interest, the empirical distribution of the extracted features is the same for all wafers. Second, we assume that the distributions within the regions of interest are vastly different from the distribution outside the regions of interest. Under these assumptions, it can be shown that for two wafermaps W_1 and W_2 with regions of interest R_1 and R_2 with empirical distributions r_1 and r_2, above equation is replaced by (Geiger, 2018, Obs. 4)

$$JSD(W_1, W_2) = \frac{|R_1|+|R_2|}{2n} H\left(\frac{|R_1|r_1+|R_2|r_2}{|R_1|+|R_2|}\right) - \frac{|R_1|}{2n}H(r_1) - \frac{|R_2|}{2n}H(r_2) + JSD(|R_1|, |R_2|),$$

where the last term is the JSD between two Bernoulli distributions with parameters R_1/n and R_2/n, respectively. This simplification has the advantage that its computation requires only the size of the regions of interest and the distributions *inside these regions*, but not the distributions on the entire wafer. It therefore fits nicely into the feature extraction pipeline proposed in (Santos et al., 2019, Sec. III.A).

Innovation

The innovation in this work is the proposal of a similarity measure to compare analog wafermaps based on process patterns. Utilizing a previously proposed feature extraction pipeline, which was shown to preserve information about patterns, we obtain an empirical distribution as a representative for each wafermap. The JSD is justified as a similarity measure because of its role in bounding classification errors (Lin, 1991, Th. 4 & 5) and at the same time having desirable properties for the analysis of analog wafermaps: It allows distinguishing wafermaps with the same pattern occurring in different sizes (Geiger, 2018, Obs. 3), and its expression simplifies if the extracted features in the “normal” wafermap region differ significantly from those in the regions of interest (Geiger, 2018, Obs. 4).

Results

To illustrate the performance of our proposed similarity measure, we conducted an experiment with a synthetic wafer test dataset comprised of 1000 analog wafermaps[1], each showing one of five commonly occurring process patterns, see Fig. 1. We utilized feature extraction pipeline from (Santos et al., 2019, Sec. III.A), with the exception that the 512-dimensional histograms were directly used, instead of performing a dimensionality reduction based on principal component analysis. The histograms were then normalized to empirical distributions.

The wafermaps were finally clustered based on these distributions. To this end, we performed hierarchical agglomerative clustering into 6, 11, and 16 clusters, using the simplified JSD as a similarity measure. In each case, the cluster assignments of the wafermaps are compared to their ground truth patterns via confusion matrices. The results are presented in Tab. 1.

One can observe that the proposed similarity measure yields good clustering results for the wafer test dataset given a sufficiently large number of clusters. Note that the proposed similarity measure discriminates patterns if the corresponding extracted features are distinct. For example, pattern 3 is a gradient over the entire wafer which can assume one out of eight different directions. The feature extraction method proposed in (Santos et al., 2019, Sec. III.A) yields features that are different for these directions, effectively splitting these wafermaps into multiple subclusters (see also (Santos et al., 2019, Fig. 1)). As one can see in Tab. 1(b), these eight directions show a stronger dissimilarity than, e.g., the features extracted from wafermaps with pattern 1 and pattern 4, which explains why the latter are put in the same cluster. (It was also observed in (Santos et al., 2019) that pattern 3 has a large intra-class variance.) This effect is even more pronounced in Tab. 1(a), where five of the six clusters are used for the different directions of pattern 3, while the remaining patterns are all put in one cluster. If the number of clusters increases, all patterns can be discriminated (see Tab. 1(c)).

[1] see https://zenodo.org/record/2542504

ABSTRACT

Acknowledgments

The work has been performed in the project Power Semiconductor and Electronics Manufacturing 4.0 (SemI40), under grant agreement No 692466. The project is co-funded by grants from Austria (BMVIT-IKT der Zukunft, FFG project no. 853338), Germany, Italy, France, Portugal and - Electronic Component Systems for European Leadership Joint Undertaking (ECSEL JU). The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Digital and Economic Affairs, and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

References

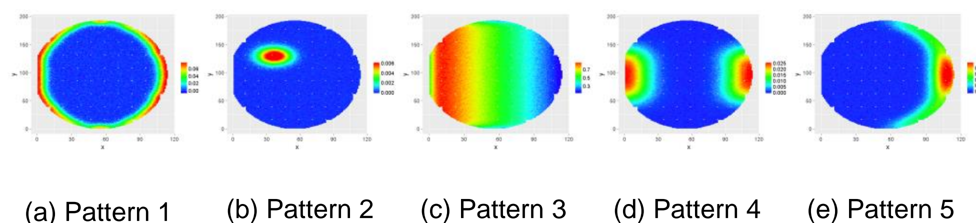
T. M. Cover & J. A. Thomas, *Elements of Information Theory*, 1st edition, New York, NY: John Wiley & Sons, Inc., 1991

B. C. Geiger, "A short note on the Jensen-Shannon divergence between simple mixture distributions," `arXiv:1812.02059 [cs.IT]`, Dec 2018.

J. F. Gómez-Lopera et al., "An Analysis of Edge Detection by Using the Jensen-Shannon Divergence", *Journal of Mathematical Imaging and Vision*, vol. 13, no. 1, pp. 35-56, Aug 2000.

J. Lin, "Divergence measures based on Shannon entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145-152, Jan 1991.

T. Santos et al., "Feature Extraction from Analog Wafermaps: A Comparison of Classical Image Processing and a Deep Generative Model," submitted to *IEEE Transactions on Semiconductor Manufacturing*.



(a) Pattern 1 (b) Pattern 2 (c) Pattern 3 (d) Pattern 4 (e) Pattern 5

Fig. 1: Five types of spatial patterns are simulated as prototypes for process pattern recognition. Pattern 1 depicts a strong area at the border of the wafer, Pattern 2 consists of a spot at arbitrary position on the wafer. Pattern 3 is a gradient with variable direction, while Pattern 4 is defined by two spots on opposite sides of the wafer. Pattern 5 is a crescent-shaped area on the right side of the wafer.

ABSTRACT

Cluster	Pattern				
	1	**2**	**3**	**4**	**5**
1	200	200		200	200
2			51		
3			24		
4			83		
5			21		
6			21		

(a) 5 clusters

Cluster	Pattern				
	1	**2**	**3**	**4**	**5**
1	200			200	
2		200			
3			27		
4			24		
5			24		
6			28		
7			21		
8			25		
9			21		
10			30		
11					200

(b) 11 clusters

Cluster	Pattern				
	1	**2**	**3**	**4**	**5**
1	200				
2		200			
3			27		
4			15		
5			10		
6			28		
7			21		
8			13		
9			9		
10			21		
11			14		
12			12		
13			14		
14			16		
15				200	
16					200

(c) 16 clusters

Tab. 1: Confusion matrices of clustering results using the presented similarity measure, compared to the ground truth of the wafermaps (= real underlying pattern). The number of clusters is limited to (a) 6 clusters, (b) 11 clusters, or (c) 16 clusters. The dataset contains five distinct patterns, each shown on 200 wafermaps in the test dataset. See text for details.