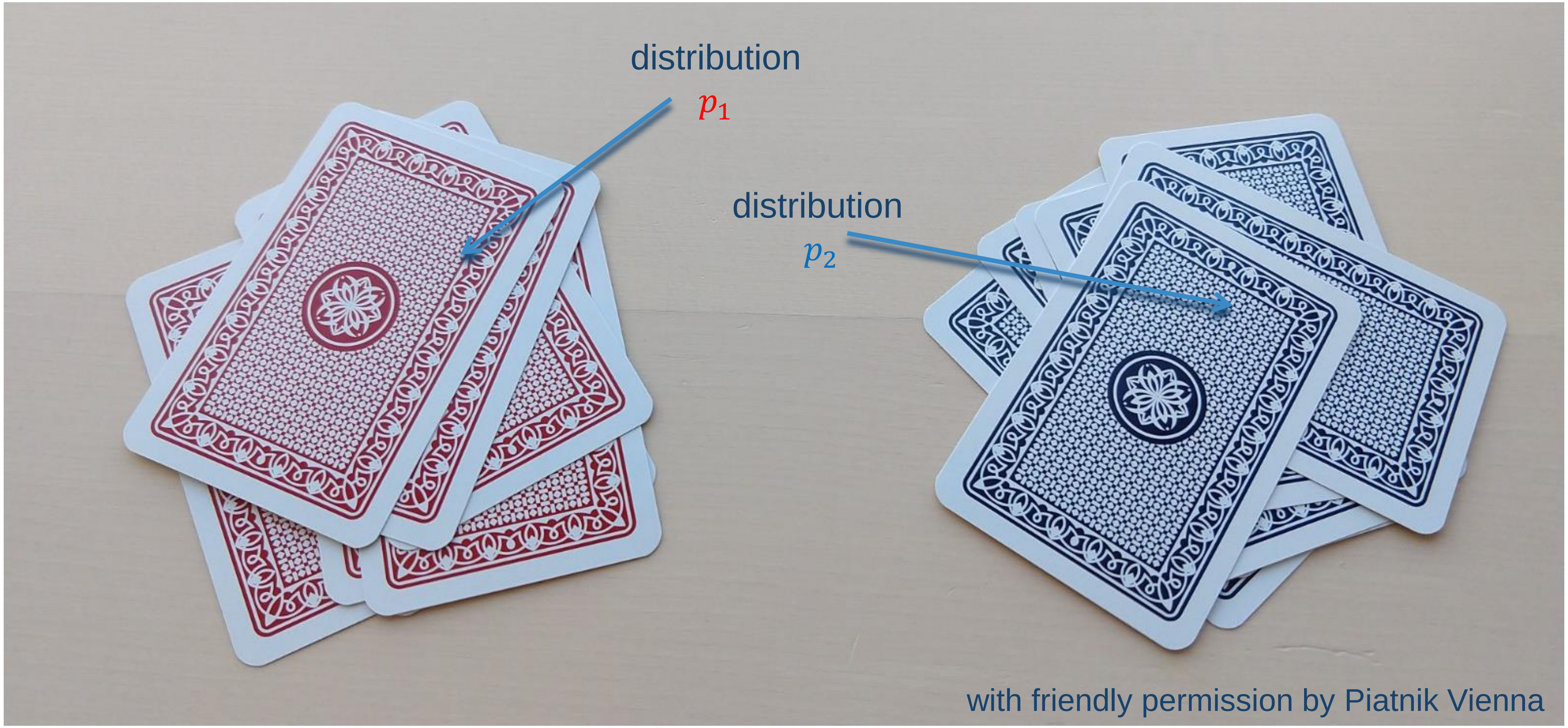
An Information-Theoretic Measure for Pattern Similarity in Analog Wafermaps

Bernhard C. Geiger,
Stefan Schrunner,
Roman Kern

The project SemI40 is co-funded by grants from Austria, Germany, Italy, France, Portugal and the ECSEL Joint Undertaking and is coordinated by Infineon Technologies Austria AG

Jensen-Shannon Divergence



Jensen-Shannon Divergence

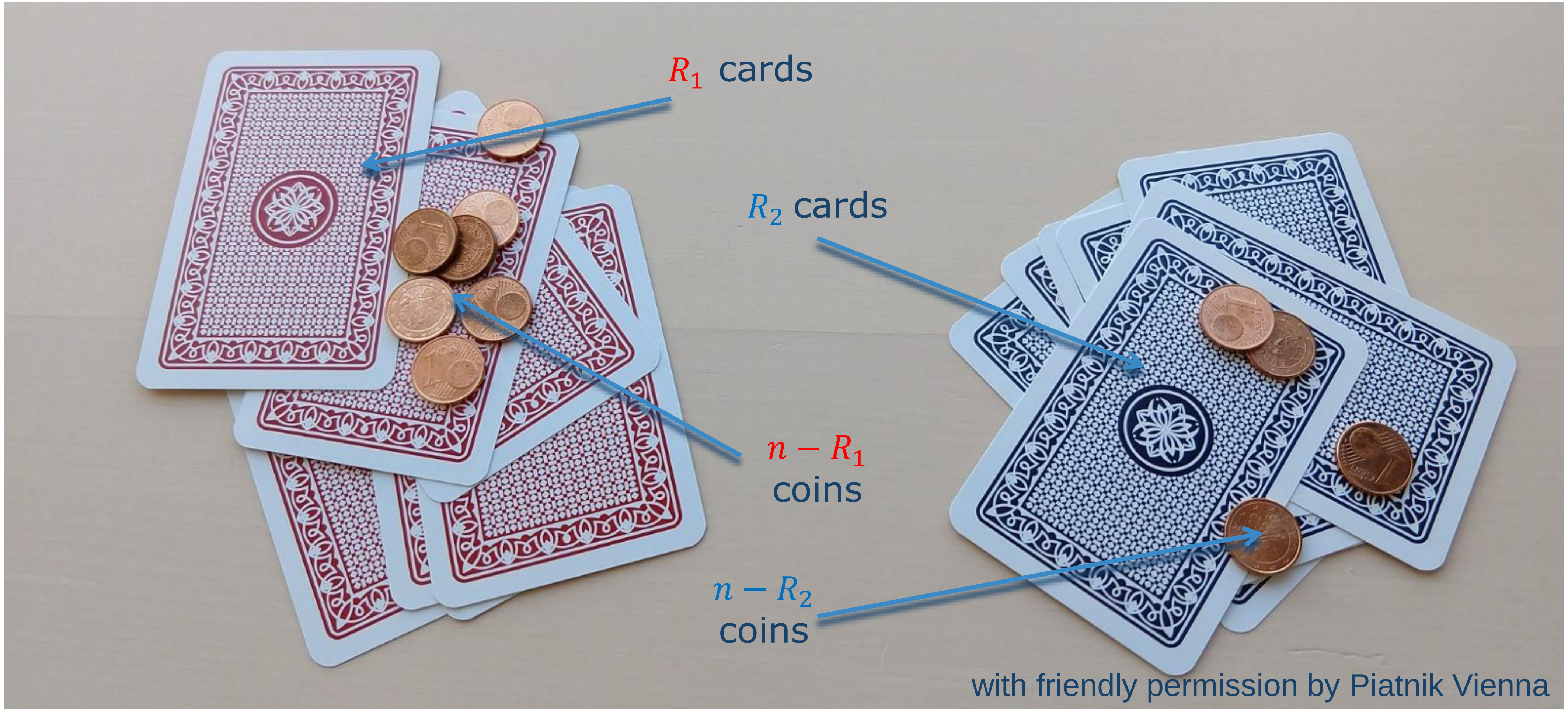
$$JSD(W_1, W_2) := H\left(\frac{p_1 + p_2}{2}\right) - 0.5H(p_1) - 0.5H(p_2)$$

provides bounds on the Bayes error:

$$\frac{1}{4}\left(1 - JSD(W_1, W_2)\right)^2 \leq P_{error} \leq \frac{1}{2}\left(1 - JSD(W_1, W_2)\right)$$

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

Jensen-Shannon Divergence



Jensen-Shannon Divergence

$$JSD(W_1, W_2) = \frac{R_1 + R_2}{2n} H\left(\frac{R_1 p_1 + R_2 p_2}{R_1 + R_2}\right) - \frac{R_1}{2n} H(p_1) - \frac{R_2}{2n} H(p_2) + JSD(r_1, r_2)$$

where r_i is Bernoulli-$\left(\frac{R_i}{n}, \frac{n-R_i}{n}\right)$.

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

Jensen-Shannon Divergence

$$JSD(W_1, W_2) = JSD(r_1, r_2)$$

with friendly permission by Piatnik Vienna

Jensen-Shannon Divergence

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

with friendly permission by Piatnik Vienna

Jensen-Shannon Divergence

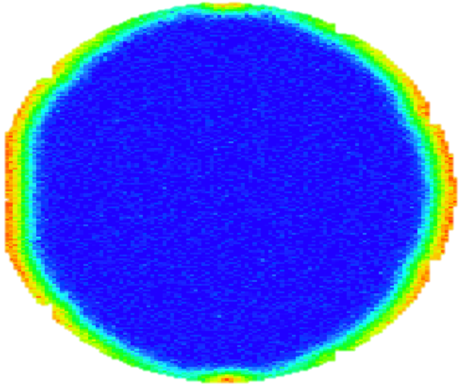
$$JSD(W_1, W_2) = 1$$

What´s the point?

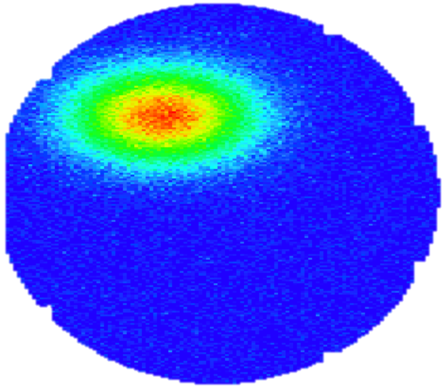
(What´s the connection to semiconductor manufacturing?)

Analog Wafermaps

Are these two wafermaps similar?



W_1



W_2

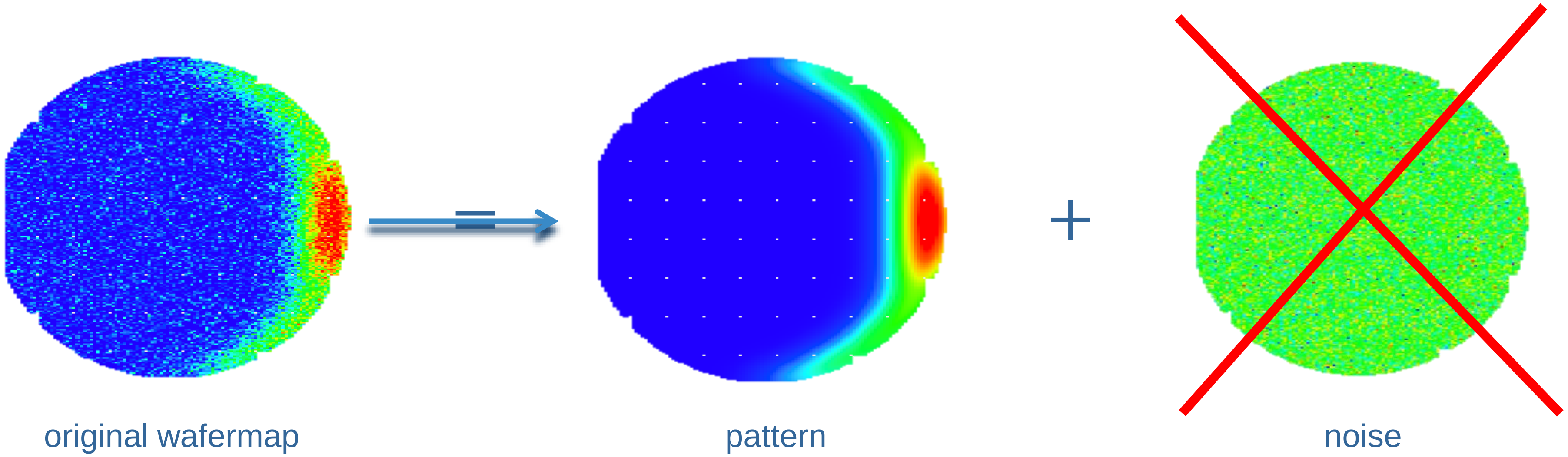
→ Clustering, Classification, Novelty Detection

Processing Pipeline

1. Denoising/Image Restoration
2. Determine „interesting“ regions on the wafer → R_1, R_2
3. Represent each device in these regions by its relation to neighbors (“binary patterns”)
4. Compute distribution of binary patterns → p_1, p_2
5. Compute similarity between wafermaps → $JSD(W_1, W_2)$

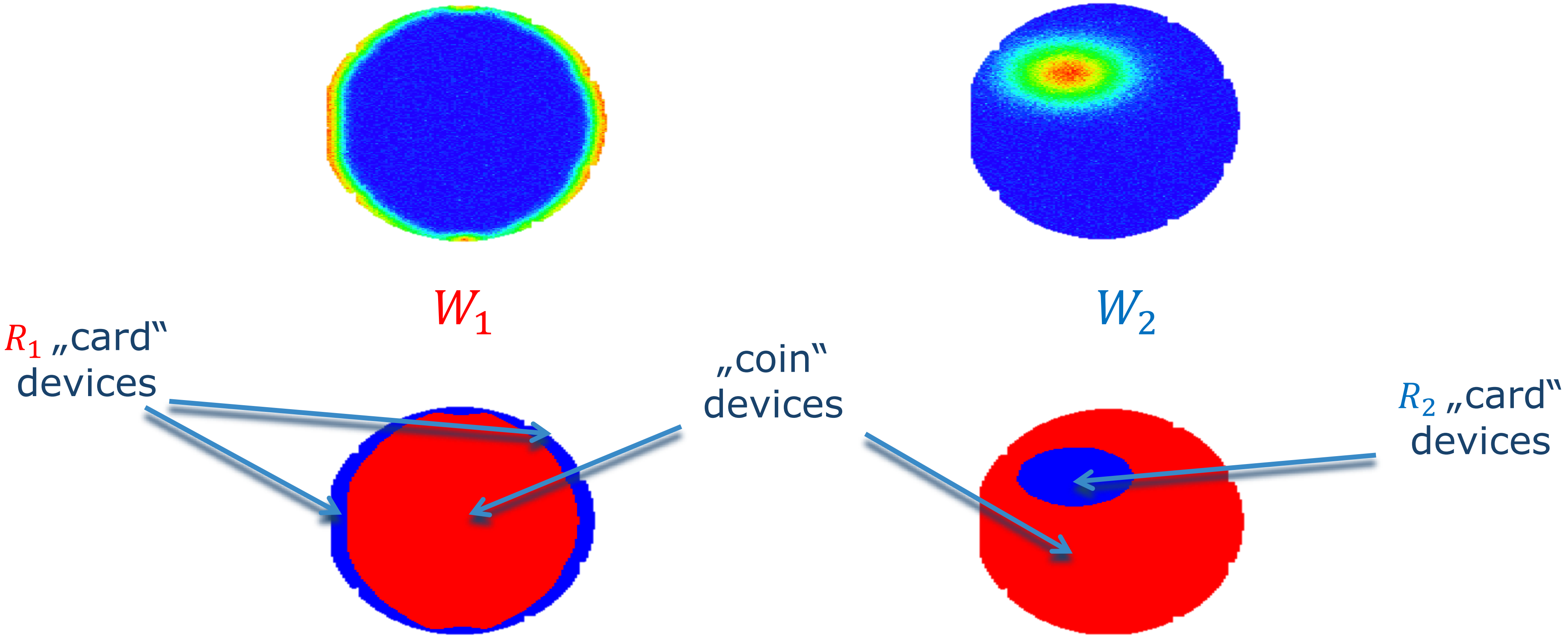
T. Santos et al., “Feature Extraction from Analog Wafermaps: A Comparison of Classical Image Processing and a Deep Generative Model,” submitted to *IEEE Transactions on Semiconductor Manufacturing.*

1. Denoising via Bayesian Inference

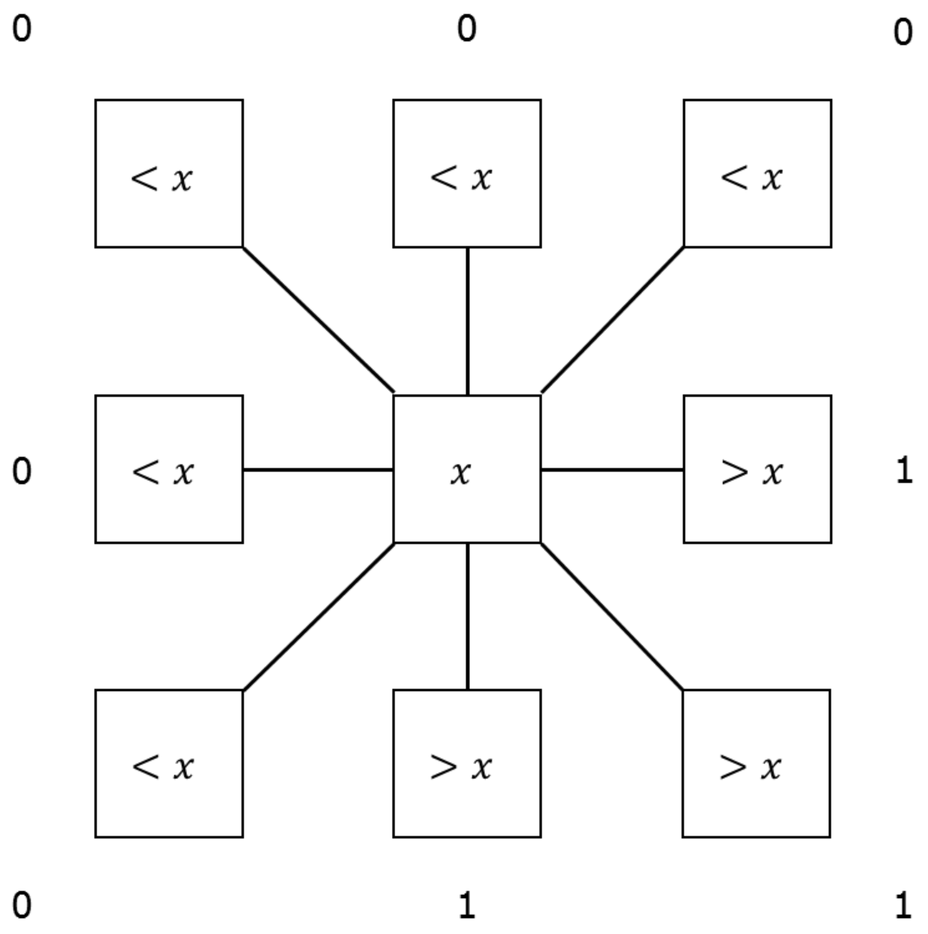


S. Schrunner et al., "Markov Random Fields for Pattern Extraction in Analog Wafer Test Data," in *Proc. Int. Conf. on Image Processing Theory, Tools and Applications,* Montreal, pp. 1-6, Nov. 2017.

2. Determine Interesting Regions

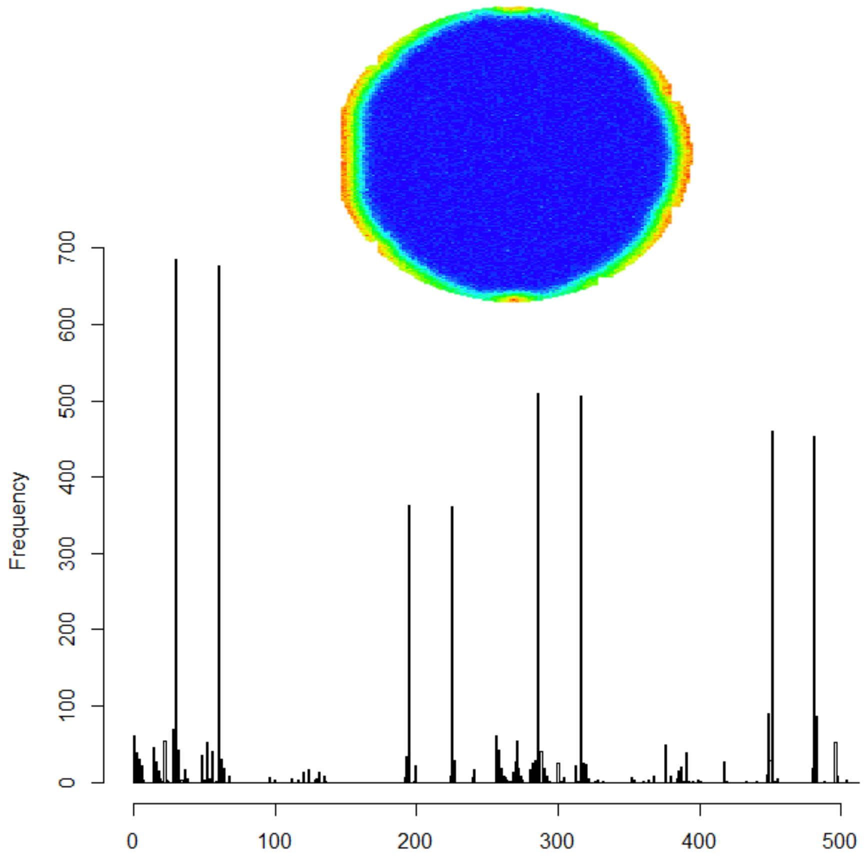


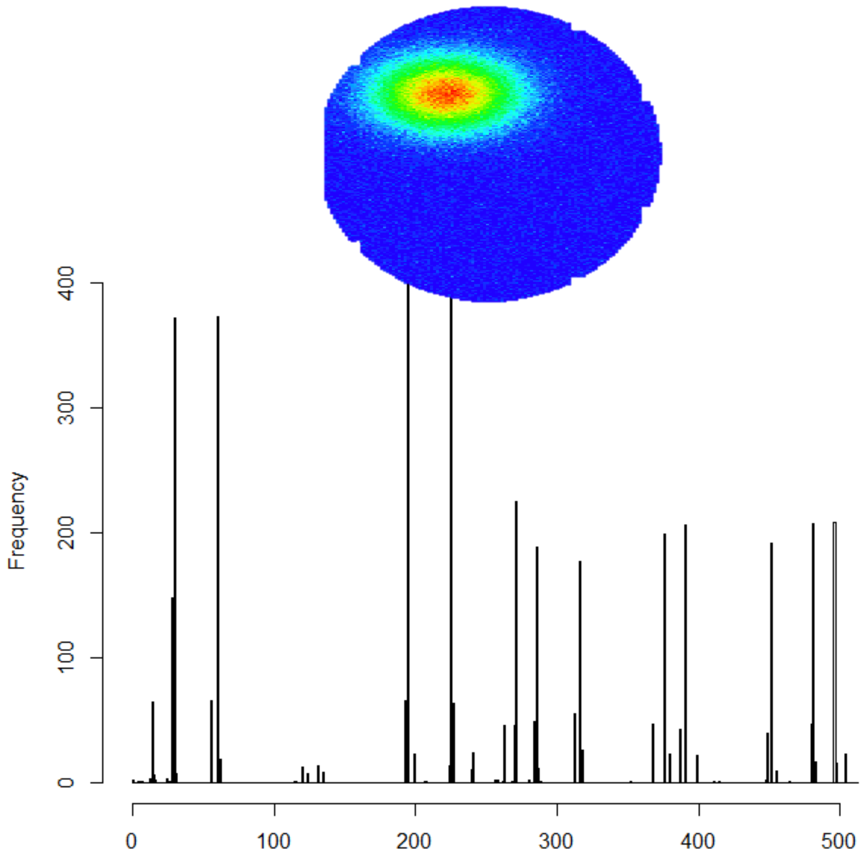
3. Local Binary Patterns



→ $(11100000)_2 = 224$

4. Empirical Distributions





5. Compute Similarity

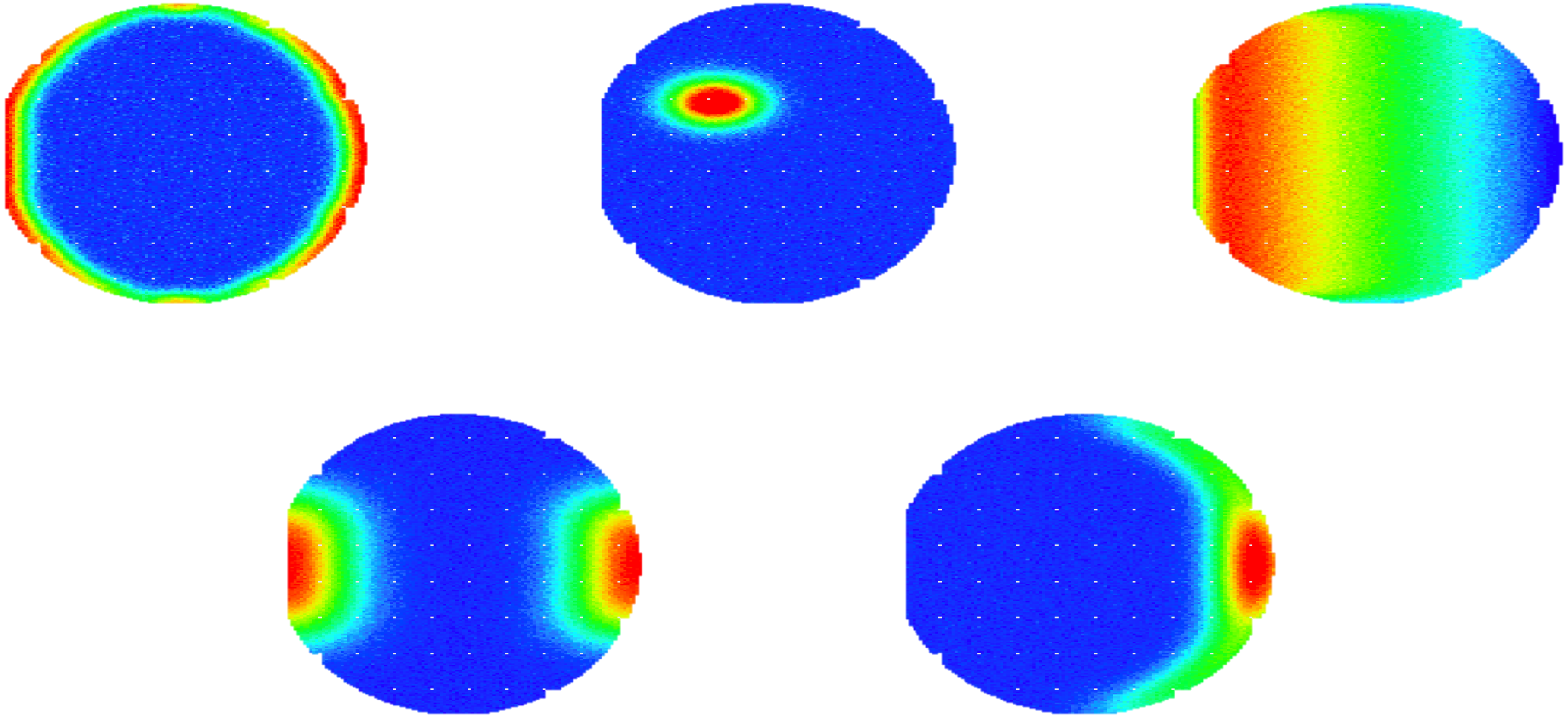
Given:

- empirical distributions $p_1, p_2, p_3, p_4, \ldots$
- sizes of interesting regions $R_1, R_2, R_3, R_4, \ldots$

Use Jensen-Shannon Divergence to compute the similarity

- ❖ between wafermaps → **Clustering**
- ❖ to prototype patterns → **Classification**
- ❖ to previously seen wafermaps → **Novelty Detection**
- ❖ etc.

Experiments



M. Pleschberger et al., “Simulated Analog Wafer Test Data for Pattern Recognition [Data set]”, 2019. https://zenodo.org/record/2542504

Confusion Matrices

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

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

Cluster	Pattern 1	Pattern 2	Pattern 3	Pattern 4	Pattern 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

Conclusion

- **Processing Pipeline**
  - pattern-specific features from analog wafermaps
  - represents wafermap via empirical distribution
- **Similarity Measure: Jensen-Shannon Divergence**
  - appropriate for histograms/distributions
  - justified via bounds on the Bayes error
- **Processing Pipeline + Similarity Measure:**
  - separates patterns from each other
  - useful for classification, novelty detection,…

Thank you!

ECSEL Joint Undertaking
Electronic Components and Systems for European Leadership