Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 52-13 (2019) 2086–2091

Adjacency-Clustering for Identifying Defect Patterns and Yield Prediction in Integrated Circuit Manufacturing ⋆

Dorit S. Hochbaum *

* *Department of Industrial Engineering and Operations Research, University of California, Berkeley, email:* (hochbaum@ieor.berkeley.edu).

Abstract: Adjacency-clustering is a new concept of capturing phenomena in the presence of spatial dependencies, or Neighborhood Effect (NE). The technique is applied here to prediction problems in the presence of NE that arise in manufacturing system monitoring, quality control and yield prediction. This work is motivated by Integrated Circuit Manufacturing (ICM) process that involves multiple steps and is exceedingly expensive. Spatial variation of parameters across each wafer, where the circuits are positioned, result from equipment or process limitations, and a circuit is likely to be defective if its neighbors on the wafer are defective as well. The existence of this Neighborhood Effect, while recognized, is not well captured in traditional modeling approaches. The challenge is to extrapolate, from given samples, the patterns of the defects and predict accurately the yield of the process. The patterns are effectively identified using adjacency-clustering that is achieved with the graph-theoretic separation-deviation model, also known as the Markov Random Field (MRF) model. The use of the technique is shown to identify the defects' patterns and provide dramatic improvements in the accuracy of yield prediction as compared to state-of-the-art methods.

Keywords: Markov Random Field, clustering, neighborhood effect, spatial variation.

1. INTRODUCTION

The process of ICM is highly complex and costly and involves hundreds of chemical or physical processing steps Yuan et al. (2011). Key processes include wafer fabrication, wafer probe, assembly or packaging and final test. The degree of manufacturing success is measured by *yield*, defined as the average ratio of the number of usable devices that pass tests after completing processes to the number of potential usable devices at the initiation of the process Kim and Kuo (1999). Accurate yield prediction is critical for managers to estimate productivity, production cost and make scheduling decisions. Moreover, yield prediction enables the detection processing problems in early production stages, which is crucial to quality improvement. A major challenge in predicting yields in the ICM process is the presence of spatial variation. Spatial variation of parameters across each wafer, where the circuits are positioned, results from equipment or process limitations, and a circuit is likely to be defective if its neighbors on the wafer are defective as well. The existence of this Neighborhood Effect (NE), while recognized, is not well captured in traditional modeling approaches. The challenge is to extrapolate, from given samples, the patterns of the defects and predict accurately the yield of the process.

We review here a new approach of adjacency-clustering (AC) for prediction problems in the presence of spatial dependencies that arise in manufacturing system monitoring, quality control and yield prediction. This approach was first presented by Hochbaum and Liu (2018). Adjacency-clustering is based on the graph-theoretic *separation-deviation* optimization problem, also known as the Markov Random Field (MRF) problem. MRF is to minimize a sum of penalty functions assigned to deviating from the priors, which are the observations derived from sampling, and penalties assigned to the violations of the spatial dependencies. This generates clusters that differentiate between segments that are characterized by varying levels of defect density while being as contiguous as possible. These clusters provide essential information about the spatial characteristics of defect patterns that enable detection of processing problems at early production stages and is used for process optimization and control. Hochbaum and Liu (2018) showed how to utilize the adjacency-clustering results to deliver highly accurate yield prediction as compared to state-of-the-art statistical techniques.

The adjacency-clustering problem can furthermore be solved optimally with an efficient algorithm, devised by Hochbaum (2001), that uses a parametric cut procedure to solve the respective MRF problem.

Spatial variation is an issue of key concern in ICM since it plays a crucial role not only for process optimization and control, but also for design of circuits that are robust to such variation. Spatial variation in ICM is manifested by the non-uniformity of defect patterns across the wafer, with e.g. more defects as the distance from the center of the wafer increases. This non-uniformity is caused by the

⋆ Research supported in part by NSF award No. CMMI-1760102.

10.1016/j.ifacol.2019.11.513

nature of the manufacturing process. Among existing techniques, Chen and Liu (2000) employed neural networks in order to recognize spatial defect patterns. Di Palma et al. (2005) tested the approach of Chen and Liu on simulated and real data set. White Jr et al. (2008) developed a procedure to detect different arrangements and shapes of defect aggregations (clusters). Recently, several recognition techniques based on support vector machines (SVM) have been tested on wafer defect data to identify different defect patterns (e.g. see Li and Huang 2009, Chao and Tong 2009, Yuan et al. 2010 and Wu et al. 2015). Ooi et al. (2013) developed an automatic defect pattern recognition system integrating feature extraction, selection and classification techniques. These methods are helpful in diagnosis but they do not explore how defect patterns can help yield prediction. The method of Hochbaum and Liu (2018) is the first to utilize defect clustering patterns in order to improve yield prediction results. Furthermore, this method is of independent interest for the purpose of identifying defect patterns with an optimization algorithm.

Most current methods for yield prediction ignore the spatial variation issue. The classical yield model assumes the number of defects on a chip follows Poisson distribution with density λ, which indicates the average number of defects on a chip and is uniform across chips and wafers. The yield is then estimated as the probability that no defects occur on a chip. In later research, the assumption of a constant λ was relaxed and instead λ was assumed to follow a certain distribution. Two popular such models are Murphy's model and Seeds' model in Murphy (1964); Seeds (1967). One shortcoming of the Poisson model is that in practice defects tend not to be randomly distributed, but rather tend to be clustered Bae et al. (2007); Hansen et al. (1997), that is, manifest the NE phenomenon.

Whereas Poisson model is widely used, it often underestimates the yield considerably for wafers with clustered defects Stapper et al. (1983); Stapper (1989). To overcome this limitation, Stapper (1983), in Stapper et al. (1983), derived a negative binomial model in which the probability that a fault occurs in a chip depends on the number of faults already on the chip. This model is equivalent to assuming that λ follows a gamma distribution Stapper et al. (1983). Albin and Friedman (1989) introduced the Neyman distribution to model defects in which both the number of clusters and the number of defects within each cluster follow Poisson distribution with different densities Albin and Friedman (1989). Although the negative binomial model and the Neyman model take the clustering of defects into account, they fail to capture the spatial information of chips. Bae et al. (2007) proposed the incorporation of positional information of chips with Poisson, negative binomial and zero-inflated Poisson regression models Bae et al. (2007). Although these regression models improve yield prediction, estimation issues remain challenging. The two main methods, Bayesian method and maximum likelihood estimation method, that are employed in the parameter estimation of regression models, have major shortcomings. The Bayesian method is based on Markov chain Monte Carlo (MCMC) and is time consuming and unstable for small-size samples. The maximum likelihood estimation method tends to provide very loose interval estimates for parameters Ghosh et al. (2006). In addition, for samples showing complicated spatial patterns, it is hard to choose covariates and set up the linear relationship in regression models.

Krueger and Montgomery Krueger and Montgomery (2014a) suggested to strengthen yield prediction models by considering temporal relationships between samples from different batches and samples over time. This adds to the "spatial" dependencies a temporal dimension and a form of correlation, or adjacency, dimension between different batches. Krueger and Montgomery (2014) introduced generalized linear mixed models (GLMM) for yield prediction that incorporate regression approaches taking into account these correlations and demonstrated that considering these additional relationships improves the quality of yield prediction. It is noted that the adjacency-clustering approach is capable of incorporating these correlations (via the separation penalty functions described in Section 2). Because of its strong accuracy results as compared to generalized linear models (GLM) the adjacency-clustering with correlations is expected to improve yield predictions even further.

2. THE SEPARATION-DEVIATION MODEL, COMPLEXITY, AND ALGORITHMS

The input to the Markov Random Field (MRF) consists of priors on a subset of the objects and spatial dependencies between pairs of spatial neighbors, implying that their estimated values should be "similar". The estimated values, or predictions, are so as to achieve a balance of two different goals: One goal is to minimize the *deviation cost* which is the cost of assigning a value different than the observation. The second goal is to minimize the *separation cost* which penalizes, for each object, the assignment of a predicted value which is different from that of the neighbor. The minimization of the combination of these two penalty functions is the *separation-deviation* problem, and the optimal values associated with the objects are the predicted true values.

The separation-deviation problem has been extensively studied in the context of image segmentation where the priors are "color" values associated with pixels. Over the past three decades, non-polynomial algorithms were devised in e.g. Blake and Zisserman (1987); Ishikawa and Geiger (1998), and later polynomial time algorithms for general convex separable penalty functions and special cases in Hochbaum (2001); Ahuja et al. (2003); Hochbaum (2013).

The MRF problem is formalized on a given graph $G = (V, E)$, where each object is represented as a node in V, and $N(i)$ is the set of neighbors of node $i \in V$. For each $j \in N(i)$ the pair of nodes $\{i, j\}$ have an edge $[i, j] \in E$ connecting them. Each node $i \in V$ that has a prior value associated with it has a deviation function $G_i()$. For d_i the observed label for node i the deviation function is of the form $G_i(x_i - d_i)$, where x_i is the output label of node i. Each edge between neighboring nodes $[i, j] \in E$ has an associated separation function $F_{ij}(x_i - x_j)$. The sets X_i are collections of discrete labels that can be assumed by variables x_i. The problem formulation is then,

$$\begin{array}{ll} \text{(MRF)} & \min \ \sum_{i\in V} G_i(x_i) + \sum_{i\in V}\sum_{j\in N(i)} F_{ij}(x_i - x_j) \\ \text{subject to} & x_i \in X_i \ \ \forall\, i \in V. \end{array}$$

The complexity of MRF depends on the form of the penalty functions. This complexity of MRF was classified according to the properties of the penalty functions in Hochbaum (2001) for feasible solution sets of consecutive integers $X_i = \{\ell_i, \ell_{i+1}, \ldots, u_i\}$, and in Hochbaum (2013) for non-consecutive values algorithm. For convex penalty functions MRF is polynomial time solvable, and for non-convex the problem is NP-hard. The cases when the deviation penalty functions are convex and the separation penalty functions are (bi-)linear (defined below) was shown by Hochbaum Hochbaum (2001) to be solvable in polynomial time using a parametric cut procedure. Furthermore, it was shown there that the complexity of the algorithm is the *fastest possible*. The case where both separation and deviation penalty functions are convex were also shown to be solvable very efficiently, and within a multiplicative log factor of fastest possible, by Ahuja et al. (2004, 2003). For non-convex deviation functions and convex separation functions the problem is solved in *pseudo-polynomial time*, that depends on the number of label values, or the range of the variables. This running time is the time required to solve a minimum cut problem on a graph with number of nodes that depends on the number of labels Ahuja et al. (2004) and it is unimprovable because the respective problem is weakly NP-hard. When both type of penalty functions are non-convex the MRF problem is NP-hard. These results, all for multi-labels MRF, are summarized in Table 1.

The notation used in Table 1 includes: $U = \max_i |X_i|$, the number of labels; $n = |V|$, the number of nodes in the graph; $m = |E|$, the number of edges in the graph; and $T(n,m)$ the complexity of the minimum s,t-cut problem on a graph with n nodes and m arcs or edges. A *bi-linear* separation function is defined to be a (two) piecewise linear function with a linear function in the range $x_i \geq x_j$ and a linear function in the range $x_j \geq x_i$, of the form:

$$F_{ij}(x_i - x_j) = \begin{cases} u_{ij}(x_i - x_j) & \text{if } x_i > x_j \\ 0 & \text{if } x_i = x_j \\ u_{ji}(x_j - x_i) & \text{if } x_i < x_j. \end{cases}$$

Table 1. Complexity of MRF problems. The references are: [AHO03] Ahuja et al. (2003); [Hoc01] Hochbaum (2001); [AHO04] Ahuja et al. (2004).

Deviation function	Separation function	Complexity	Ref.
Convex	Convex	$O(mn \log \frac{n^2}{m} \log nU)$	[AHO03]
Convex	Bilinear	$O(mn \log \frac{n^2}{m} + n \log U)$	[Hoc01]
Quadratic	Bilinear	$O(mn \log \frac{n^2}{m})$	[Hoc01]
General	convex	$O(mnU^2 \log \frac{n^2 U}{m})$	[AHO04]
Linear	Nonlinear	NP-hard	[Hoc01]

It is noted that the complexities of the algorithms of Hochbaum (2001) and Ahuja et al. (2004) stated in Table 1 for convex-bilinear, quadratic-bilinear and nonlinear-convex, are provably fastest possible.

3. MRF FOR YIELD PREDICTION AND EXPERIMENTAL RESULTS

Our yield prediction approach works by first partitioning the wafer into clusters incorporating the neighborhood effect so that each cluster consists of chips that share the same estimated likelihood – yield level. In different clusters the yield level differs. Once the clusters have been identified, the overall wafer yield level is a weighted (by size) combination of the yield levels in the clusters.

Let $(d_1, d_2, \ldots, d_n)$ be observed number of defects on a wafer map with n chips, where d_i represents the observed number of defects at the i-th chip. This observed number is not necessarily provided for all chips on the map. If unavailable for position i, d_i is considered a "don't-care" value and there is no penalty term for deviating from that value. The wafer map is represented as a graph $G = (V, E)$ where nodes represent chips and edges define the neighborhood relationship, here selected as the 4-neighbor system (rook-move neighborhood). The goal is to associate with each chip a label x_i that indicates the cluster it belongs to. For X_i the list of labels that characterize the clusters, general, $x_i \in X_i$, we select the set of values $\{0, 1, \ldots, k\}$, where k is a parameter set by the user. The choice of k implies that there are $(k+1)$ potential labels that characterize the yield level of each chip. For instance, if $k = 2$, a wafer is partitioned into three types of clusters: non-defective ($x_i = 0$), medium defective ($x_i = 1$) and highly defective ($x_i = 2$). Another possible choice, for a wafer with number of fatal defects that vary, say, between 0 and 5, to let $X_i = \{0, 2, 5\}$ with the same interpretation, non-defective, medium and highly defective.

We select here uniform quadratic deviation functions and absolute value separation functions: The deviation function $G_i(x_i) = (d_i - x_i)^2$ is commonly used in image restoration Boykov et al. (1999). For the separation penalties we let $F_{ij}(x_i - x_j) = \alpha \cdot |x_i - x_j|$. The factor α is a scalar that is used to balance the separation versus the deviation penalties. The selection of α is selected here to be uniform for all pairs as there is no predominant difference between pairs. The problem is then,

$$\begin{array}{ll} \min & \sum_{i\in V} (x_i - d_i)^2 + \sum_{[i,j]\in E} \alpha |x_i - x_j| \\ \text{subject to} & x_i \in \{0, 1, \ldots, k\} \ \ \forall\, i \in V. \end{array}$$

To solve the problem we apply the parametric cut procedure of Hochbaum (2001). The adjacency-clustering procedure was applied to four wafer maps: One is from Tygai and Bayoumi (1994) Tyagi et al. (1994) and three from Yuan et al. (2011) Yuan et al. (2011).

We first present the effect of the different parameters on the resulting clusters for the first wafer map, in Figure 1. The effect of increasing the number of cluster labels k for k (=1,2,3) (going down the rows of images), the clusters corresponding to positive values of the label are becoming more differentiated into small contiguous groups. Each cluster is indicated by a different color with redder color implying higher congregation of defects. The tests were run for four different values of α (=0.1,0.5,1,2) across columns from left to right respectively. The effect of higher values of α is manifested in the tendency to create larger, more

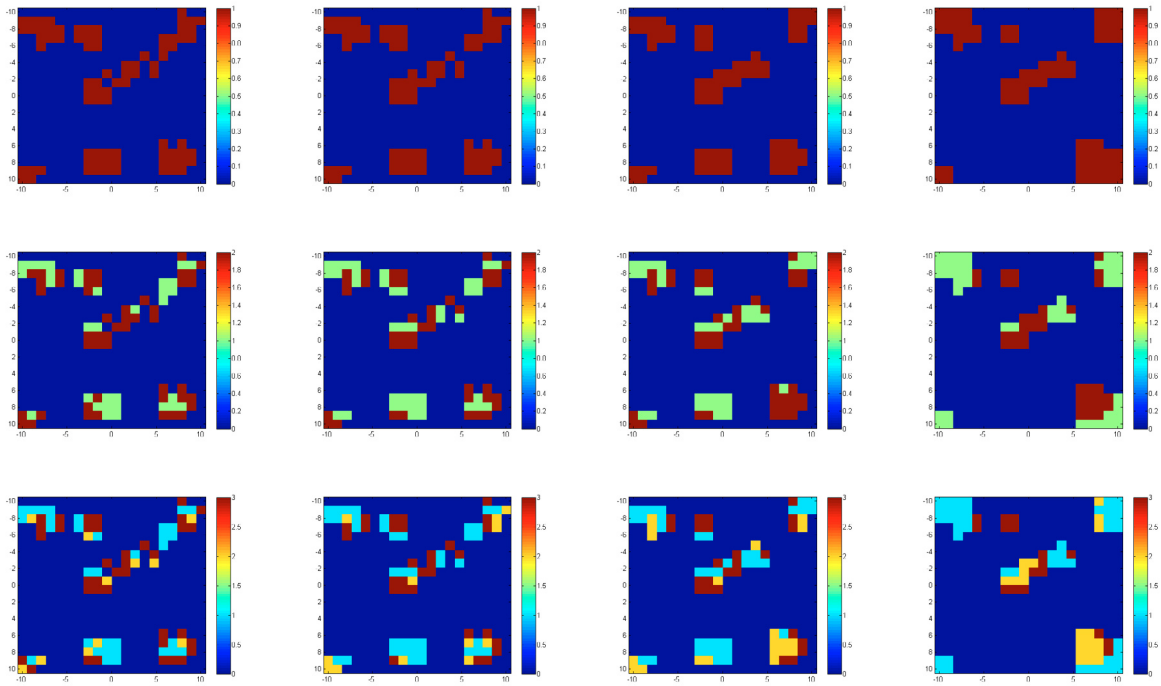

Fig. 1. Separation-deviation clustering results for wafer map 1 (source Hochbaum and Liu (2018).)

cohesive, clusters. This is due to the heavier emphasis on the separation penalty as the value of α increases. The pattern observed here, for wafer map 1, is that high defect density clusters are positioned near the center and four corners of the wafer, which may imply possible manufacturing problems.

3.1 Yield Prediction

For the attained clusters, the overall yield is determined as follows: For cluster $\mathcal{C}_j$, λ_j is the average number of defects in the cluster and the yield is estimated as $\exp(-\lambda_j)$. That is, for each cluster we used the Poisson yield model for the cluster's yield prediction. We describe below the use of other yield models for each cluster.

The total yield is determined as the weighted average of cluster yields, $\sum_{j=1}^{\ell} |\mathcal{C}_j| \exp(-\lambda_j) / \sum_{j=1}^{\ell} |\mathcal{C}_j|$. The standard performance measure is the *relative absolute bias* (Yuan et al. (2011)), $\frac{|\text{True yield} - \text{Estimated yield}|}{\text{True yield}}$.

From our empirical results, and use of machine learning techniques for selecting the parameter values (not described here for lack of space), the selection of $\alpha = 1$ dominates other values, and selecting the value $k = 2$ yields fairly close results to those for the best selected value. With these parameters we implemented the adjacency-clustering approach with the Poisson model for each cluster. The results of the comparison of the relative absolute bias of the adjacency-clustering model with the classical Poisson model and Poisson regression model are presented in Table 2. The results indicate that Poisson regression is more accurate than Poisson, and the separation-deviation model is significantly more accurate than both of them.

3.2 Adjacency-Clustering with Different Yield models

In addition to the Poisson yield model, the negative binomial model is also widely used in yield prediction. Compared with Poisson model, it is less likely to underestimate the yield (Kim 2011). Following negative binomial model, the yield for cluster j is given by $\hat{y}_j = (1+\lambda_j/\gamma_j)^{-\gamma_j}$, where γ_j is called the cluster parameter. There are multiple ways of determining γ_j (see Cunningham 1990 for details), and we adopt the method of moments with $\gamma_j = \frac{\lambda_j^2}{\sigma_j^2 - \lambda_j}$ where σ_j^2 is the variance of the number of defects per chip for the cluster. Three different yield models applied to the clusters of adjacency-clustering are used here: (1) AC-NB model: negative binomial yield model is fitted to each cluster; (2) AC-NBP model: negative binomial yield model is fitted to non-defective clusters (cluster with "0"s) while Poisson yield model is fitted to defective clusters (of label > 1); (3) AC-PNB model: Poisson yield model is fitted to the non-defective cluster while negative binomial yield model is applied to defective clusters. These were tested on the four wafers for different combinations of α and k. We select the parameter values that yield the lowest prediction errors (AC-NB: $\alpha = 0.7$, $k = 3$; AC-NBP: $\alpha = 0.6$, $k = 1$; AC-PNB: $\alpha = 0.7$, $k = 3$). (Experimental results for the choice of these parameter values are provided in Hochbaum and Liu (2018).) It should be noted that the choice of $\alpha = 1$ and $k = 2$ achieves similar results to the above parameters with average gaps of 0.0036, 0.0019 and 0.0083 for AC-NB, AC-NBP and AC-PNB respectively.

The prediction results of these three AC models are compared with negative binomial model and negative binomial regression model in Table 3. In negative binomial regression model,the same covariates are selected as in Poisson

Table 2. Yield predictionfor AC with Poisson model and Poisson regression model (source Hochbaum and Liu (2018).)

Model	Data set	1	2	3	4
	True yield	79.50%	84.40%	89.90%	79.30%
Poisson model	Yield prediction	52.33%	74.90%	87.90%	72.20%
	Relative Absolute Bias	34.18%	11.26%	2.22%	8.95%
Poisson regression model	Yield prediction	55.30%	79.40%	88.40%	72.60%
	Relative Absolute Bias	30.44%	5.92%	1.67%	8.45%
Adjacency-clustering model	Yield prediction	81.09%	82.84%	89.38%	78.49%
($\alpha = 1, k = 2$)	Relative Absolute Bias	**2.00**%	**1.85**%	**0.58**%	**1.02**%

regression model, and the coefficients are estimated using maximum likelihood method, which is implemented in *glm.nb()* in R.

The results in Table 3 indicate that AC-NB is the best model to use uniformly. Specifically, AC-NB model outperforms other models for wafer 3 and wafer 4, AC-NBP model yields the best result for wafer 1 and AC-PNB model gives the best result for wafer 2. Compared with the negative binomial model, AC-NB model improves the prediction result by a factor between 2 and 14. Compared with negative binomial regression model, the error of AC-NB model is lower on wafers 2 and 4 and about the same for wafer 3, and a bit worse for wafer 1. The reason why in some of the cases AC-PNB and AC-NBP perform better than AC-NB is that combining different yield models works better in cases of unstable manufacturing processes that render different defect behaviors in different areas on the wafer. Among all these, AC-NB model is uniformly the most robust and therefore it is a recommended method for yield prediction.

4. CONCLUSION

We present here the adjacency-clustering method for yield prediction which is unique in that it generates first defect-pattern clusters characterized by the level of defects expected within the clusters. The contiguity of these clusters, and the respective neighborhood effect, is a trade-off parameter incorporated into the optimization problem of MRF used to identify such clusters. The generated clusters are then used by the yield prediction task, by generating yield prediction for each homogeneous cluster and then combining the results. The power of adjacency-clustering for yield prediction was demonstrated by Hochbaum and Liu (2018) via an extensive empirical study for real and simulated data sets in which significant improvements in yield prediction were attained. For simulated data sets the results indicated that the method provides significant improvement in yield prediction as compared to existing statistical techniques. It was further observed that as the neighboring effect increases in the simulated data sets, the advantage provided by AC gets magnified.

Future work is to include longitudinal correlations between and within batches of samples as introduced by Krueger and Montgomery (2014b), who showed that considering these additional relationships improves the quality of yield prediction. Still, their approach ignores the neighborhood effect. We expect that the incorporation of these correlations in the adjacency-clustering approach will result in yet further improvements in yield prediction.

Finally, the generation of the defect-pattern clusters with adjacency-clustering has not yet been explored as to its potential in identifying issues in the manufacturing process. Since the generation of the clusters is done with an optimization process that takes into consideration the confidence in the priors and the strength of the neighborhood effect in different areas (through a judicious selection of the deviation and separation functions) it has the potential of improving on the state-of-the-art in identifying spatial variations and defect patterns.

REFERENCES

Ahuja, R.K., Hochbaum, D.S., and Orlin, J.B. (2003). Solving the convex cost integer dual network flow problem. *Management Science*, 49(7), 950–964.

Ahuja, R.K., Hochbaum, D.S., and Orlin, J.B. (2004). A cut-based algorithm for the nonlinear dual of the minimum cost network flow problem. *Algorithmica*, 39(3), 189–208.

Albin, S.L. and Friedman, D.J. (1989). The impact of clustered defect distributions in ic fabrication. *Management Science*, 35(9), 1066–1078.

Bae, S.J., Hwang, J.Y., and Kuo, W. (2007). Yield prediction via spatial modeling of clustered defect counts across a wafer map. *IIE Transactions*, 39(12), 1073–1083.

Blake, A. and Zisserman, A. (1987). *Visual reconstruction*. MIT press.

Boykov, Y., Veksler, O., and Zabih, R. (1999). Fast approximate energy minimization via graph cuts. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 1, 377–384. IEEE.

Chao, L.C. and Tong, L.I. (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, F.L. and Liu, S.F. (2000). A neural-network approach to recognize defect spatial pattern in semiconductor fabrication. *Semiconductor Manufacturing, IEEE Transactions on*, 13(3), 366–373.

Cunningham, J.A. (1990). The use and evaluation of yield models in integrated circuit manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 3(2), 60–71.

Table 3. Yield prediction for: AC-NB, AC-PNB, AC-NBP , negative binomial and negative binomial regression models. (source Hochbaum and Liu (2018).)

Wafer Map	1	2	3	4
True yield	79.50%	84.36%	89.85%	79.28%
AC-NB ($\alpha = 0.7$, $k = 3$)	79.82%	84.12%	89.94%	79.19%
Relative Absolute Bias	0.40%	0.27%	**0.10**%	**0.12**%
AC-NBP ($\alpha = 0.6$, $k = 1$)	79.33%	84.47%	90.54%	80.06%
Relative Absolute Bias	**0.22**%	0.14%	0.76%	0.98%
AC-PNB ($\alpha = 0.7$, $k = 3$)	79.83%	84.34%	90.18%	79.78%
Relative Absolute Bias	0.41%	**0.02**%	0.37%	0.63%
Negative binomial model	76.31%	83.09%	89.71%	78.07%
Relative Absolute Bias	4.01%	1.51%	0.16%	1.53%
Negative binomial regression model	79.28%	84.12%	89.76%	79.07%
Relative Absolute Bias	0.28%	0.28%	**0.10**%	0.26%

Di Palma, F., De Nicolao, G., Miraglia, G., Pasquinetti, E., and Piccinini, F. (2005). Unsupervised spatial pattern classification of electrical-wafer-sorting maps in semiconductor manufacturing. *Pattern recognition letters*, 26(12), 1857–1865.

Ghosh, S.K., Mukhopadhyay, P., and Lu, J.C.J. (2006). Bayesian analysis of zero-inflated regression models. *Journal of Statistical planning and Inference*, 136(4), 1360–1375.

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

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

Hochbaum, D.S. (2013). Multi-label markov random fields as an efficient and effective tool for image segmentation, total variations and regularization. *Numerical Mathematics: Theory, Methods and Applications*, 6(1), 169–198.

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

Ishikawa, H. and Geiger, D. (1998). Segmentation by grouping junctions. In *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, 125–131. IEEE.

Kim, K.O. (2011). Burn-in considering yield loss and reliability gain for integrated circuits. *European Journal of Operational Research*, 212(2), 337–344.

Kim, T. and Kuo, W. (1999). Modeling manufacturing yield and reliability. *IEEE Trans. Semicond. Manuf*, 12(4), 485–492.

Krueger, D. and Montgomery, D. (2014a). Modeling and analyzing semiconductor yield with generalized linear mixed models. *Applied Stochastic Models in Business and Industry*, 30(6), 691–707.

Krueger, D. and Montgomery, D. (2014b). Modeling and analyzing semiconductor yield with generalized linear mixed models. *Applied Stochastic Models in Business and Industry*, 30(6), 691–707.

Li, T.S. and Huang, C.L. (2009). Defect spatial pattern recognition using a hybrid som–svm approach in semiconductor manufacturing. *Expert Systems with Applications*, 36(1), 374–385.

Murphy, B.T. (1964). Cost-size optima of monolithic integrated circuits. *Proceedings of the IEEE*, 52(12), 1537–1545.

Ooi, M.P.L., Sok, H.K., Kuang, Y.C., Demidenko, S., and Chan, C. (2013). Defect cluster recognition system for fabricated semiconductor wafers. *Engineering Applications of Artificial Intelligence*, 26(3), 1029–1043.

Seeds, R. (1967). Yield and cost analysis of bipolar lsi. In *Electron Devices Meeting, 1967 International*, volume 13, 12–12. IEEE.

Stapper, C.H. (1989). Large-area fault clusters and fault tolerance in vlsi circuits. *IBM Journal of research and Development*, 33(2), 162–173.

Stapper, C.H., Armstrong, F.M., and Saji, K. (1983). Integrated circuit yield statistics. *Proceedings of the IEEE*, 71(4), 453–470.

Tyagi, A., Bayoumi, M., et al. (1994). The nature of defect patterns on integrated-circuit wafer maps. *IEEE Trans. Rel.*, 43(1), 22–29.

White Jr, K.P., Kundu, B., and Mastrangelo, C.M. (2008). Classification of defect clusters on semiconductor wafers via the hough transformation. *IEEE Trans. Semicond. Manuf*, 21(2), 272–278.

Wu, M.J., Jang, J.S.R., and Chen, J.L. (2015). Wafer map failure pattern recognition and similarity ranking for large-scale data sets. *Semiconductor Manufacturing, IEEE Transactions on*, 28(1), 1–12.

Yuan, T., Bae, S.J., and Park, J.I. (2010). Bayesian spatial defect pattern recognition in semiconductor fabrication using support vector clustering. *The International Journal of Advanced Manufacturing Technology*, 51(5-8), 671–683.

Yuan, T., Ramadan, S.Z., and Bae, S.J. (2011). Yield prediction for integrated circuits manufacturing through hierarchical bayesian modeling of spatial defects. *IEEE Trans. Rel.*, 60(4), 729–741.