SPATIAL STOCHASTIC PROCESSES FOR YIELD AND RELIABILITY MANAGEMENT WITH APPLICATIONS TO NANO ELECTRONICS

A Dissertation

by

JUNG YOON HWANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2004

Major Subject: Industrial Engineering

SPATIAL STOCHASTIC PROCESSES

FOR YIELD AND RELIABILITY MANAGEMENT

WITH APPLICATIONS TO NANO ELECTRONICS

A Dissertation

by

JUNG YOON HWANG

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Way Kuo (Chair of Committee)	
Alberto Garcia-Diaz (Member)	Daniel W. Apley (Member)
Thomas E. Wehrly (Member)	Brett A. Peters (Head of Department)

December 2004

Major Subject: Industrial Engineering

ABSTRACT

Spatial Stochastic Processes for Yield and Reliability Management
with Applications to Nano Electronics. (December 2004)
Jung Yoon Hwang, B.A., Korea Military Academy;
M.S., Texas A&M University
Chair of Advisory Committee: Dr. Way Kuo

This study uses the spatial features of defects on the wafers to examine the detection and control of process variation in semiconductor fabrication. It applies spatial stochastic process to semiconductor yield modeling and the extrinsic reliability estimation model. New yield models of integrated circuits based on the spatial point process are established. The defect density which varies according to location on the wafer is modeled by the spatial nonhomogeneous Poisson process. And, in order to capture the variations in defect patterns between wafers, a random coefficient model and model-based clustering are applied. Model-based clustering is also applied to the fabrication process control for detecting these defect clusters that are generated by assignable causes. An extrinsic reliability model using defect data and a statistical defect growth model are developed based on the new yield model.

ACKNOWLEDGMENTS

Special thanks to my advisor, Professor Way Kuo for his patience, support, guidance and continuous attention. His vision of research trends and his timely and accurate guidance led this study to the right destination. I thank my advisory committee Professor Alberto Garcia-Diaz, Professor Daniel W. Apley, and Professor Thomas E. Wehrly for their support.

I would like to thank my parents and parents-in-law. I am grateful to my wife, Sang Nam Park, who made this study possible with her patience, support and love, and to my son, Sung Joon Hwang.

This study is supported by NSF Projects DMI-0429176, DMI-0300032, and DMI-0243409.

TABLE OF CONTENTS

CHAPTER Page

I INTRODUCTION 1

II YIELD AND YIELD LEARNING 7

II.1. Yield 8
II.2. Defects and defect density 8
II.3. Defect size distribution 9
II.4. Critical area 11
II.5. Yield models 14
II.6. Limitations of existing yield models 18
II.7. Defect monitoring and yield learning 20
II.8. Summary 21

III RELIABILITY 23

III.1. Reliability concepts 23
III.2. Burn-in 25
III.3. Yield-reliability relation models 27
III.4. Summary 30

IV YIELD MODELING OF INTEGRATED CIRCUITS BASED ON SPATIAL NONHOMOGENEOUS POISSON PROCESS 31

IV.1. Introduction 31
IV.2. Yield model review 33
IV.3. Spatial nonhomogeneous Poisson process 37
IV.4. Yield models 39
IV.4.1. Point defect yield model 39
IV.4.2. General defect yield model 41
IV.5. Intensity function 42
IV.5.1. Estimation 42
IV.5.2. Construction 44
IV.6. Simulation 48
IV.6.1. Simulation setup 48
IV.6.2. Simulation results 51
IV.7. Conclusion 54

CHAPTER Page

V MODEL-BASED CLUSTERING FOR INTEGRATED CIRCUIT YIELD ENHANCEMENT 55
V.1. Introduction 55
V.2. Model-based clustering 60
V.2.1. Probability models 60
V.2.2. EM algorithm for clustering 63
V.2.3. CEM algorithm 64
V.2.4. Number of clusters 65
V.2.5. Principal curve 66
V.3. The clustering strategy 68
V.3.1. Global defect pattern and distribution 68
V.3.2. Local defect patterns 69
V.3.3. Statistical clustering algorithm 70
V.4. Simulated defects 71
V.5. Conclusion 72

VI EXTRINSIC RELIABILITY ESTIMATION FROM YIELD INFORMATION 75
VI.1. Introduction 75
VI.2. Competing risk model 78
VI.3. Yield model based on spatial NHPP 80
VI.3.1. Spatial nonhomogeneous Poisson process 80
VI.3.2. Yield model 81
VI.4. Defect growth model 84
VI.4.1. Mixture defect size distribution 84
VI.4.2. Statistical defect growth model 85
VI.5. Extrinsic reliability model 86
VI.6. Failure rate function 89
VI.7. Conclusion and future research 91

VII CONCLUDING REMARKS 92

REFERENCES 95

VITA 102

LIST OF FIGURES

FIGURE		Page
1	Defect size distribution	10
2	Defects and faults	12
3	Chip with single conduction line	12
4	Defect density distributions: half Gaussian, exponential, triangular, uniform and gamma distribution($\alpha = 2$)	18
5	Yield models	19
6	Bathtub-shaped failure rate curve	26
7	Simulated defects	49
8	Critical area	51
9	Simulated defect patterns	58
10	Principal components	62
11	Bayesian information criteria	66
12	Principal curve	68
13	Clustering results of simulated defects	74
14	Yield and reliability relation with various k values	88
15	Extrinsic failure rate curve in arbitrary units with $\eta(t, e) = \frac{e}{a} \log \frac{t}{b}$	90

LIST OF TABLES

TABLE Page

1 Compound Poisson yield models 19

2 Simulation with various intensity functions 52

3 Simulation with or without clustering algorithm 52

4 Simulation of the general defect yield model with various defect size distributions and critical area functions 53

CHAPTER I

INTRODUCTION

The importance of nanotechnology is growing in many fields, such as medicine, semiconductor fabrication, materials, etc. Although this emerging technology has tremendous potential for changing human life in the near future, there is a barrier to overcome before the mass production of nano products is feasible. It is known that both the yield and reliability of nano products are very low [1]. One of the unique characteristics of nano products is feature sizes. The feature sizes are so small that atomic or molecular level precision is required. In the nano world where quantum physics and chemical reactions come into play, small variations caused by human factors, parameter settings of equipment or environmental factors, which have been ignored in traditional production, may lead to big changes in the performance of the products. Nanotechnology is beyond the scope of classic physics where the manipulation of parameters is allowed, and the objects in the process are so small that they are susceptible to numerous variations. In order to improve the yield and reliability of nano products, methods of controlling and reducing the variations are needed, and detecting these variations is the starting point for improving nano product yield and reliability. This study investigates methods for monitoring and detecting variations in the semiconductor fabrication process.

In semiconductor fabrication, which has more than two hundred process steps each with more than ten key process measures, tracing the process variation by monitoring these measures is not well developed nor efficient [2]. Summary measures like

---

This dissertation follows the style of the *IEEE Transactions on Reliability*.

defect counts, defect density or yield, are typically monitored instead. Variations in the fabrication process are the source of yield loss. Thus, the magnitude of variation can be partially modeled by yield models. However, yield is a summary measure of process variations and is not sensitive enough to capture all of the variations. The parameters of yield models, such as defect density distributions in compound Poisson yield models, also provide partial information about process variations since variation is modeled numerically by the defect density distributions without providing any information about defect density variation by location.

Defects are the main source of both yield loss and extrinsic reliability loss in semiconductors. Since defects are generated from variations in the fabrication processes, traces of the variations are contained in the defects. Hence, finding and removing the variations can be achieved by examining the defects on the wafers. If the defects were uniformly distributed, i.e., defects generated from a homogeneous Poisson process, yield modeling and process control would be much easier. However, defect density is known to vary within wafers, between wafers and between lots. The spatial and temporal factors of process variations are the cause of the nonhomogeneity in defect densities. Thus, useful information about yield loss and extrinsic reliability can be retrieved from the defect pattern. Up until now, the study of defect density variation in yield modeling and the study of defect patterns for process control have been conducted separately in different ways since yield models do not have capability to model the spatial defect pattern on the wafer.

Yield is one of the very important measures of the semiconductor fabrication process. Since it is a measure of the success of a manufacturing process and is closely related to the profitability, it has been a major concern of the semiconductor industry for many years. In the semiconductor industry where the manufacturers can sell all of the products they produce, yield is an even more important factor than in any

other industry. Yield is estimated and predicted by yield models. In addition to estimating or predicting yield, yield models are used in the performance evaluation of the fabrication process, including investment decisions about new products and production cost estimations. Thus, building and utilizing an accurate yield model is essential to the success of the semiconductor manufacturer. However, accurate yield modeling is still a challenging task. In existing compound Poisson yield models, the defect density variation is modeled by the defect density distribution, which cannot describe the spatial variation of the defect density. The defect density distributions, estimated from different layers of integrated circuits or different kinds of defects, cannot be incorporated to get an overall defect density distribution. To overcome this limitation of existing yield models, the spatial point process is adopted because it models the spatial variation of defect patterns very well.

Considering the tremendous expense of semiconductor fabrication facilities and the rapid advance of the technology, manufacturers have to improve their yield on an appropriate time schedule. The yield improvement process is called yield learning. Yield learning represents the dynamic aspect of the procedure that is strongly governed by time. Rapid yield learning is critical for the economic success of a semiconductor manufacturing process. Several months difference in a yield improvement schedule can result in a net profit loss of billions of dollars [3]. Yield learning is a continuous process that discovers the causes of yield loss and removes them. Finding the causes of yield loss is done by examining defects on the wafers. Subjective decisions are made based on the results of defect inspection of the size, color or shape of the defects. Considering the huge size of wafer defect data, it is too expensive to examine all of the defects [4]. Thus, inspection efforts are focused on defects generated by assignable causes. The assignable causes generate multiple defects which leave specific defect patterns on the wafer. These spatial defect patterns

can be utilized to increase the efficiency of defect inspection for root-cause analyses. Using model-based clustering, the defects from assignable causes are identified by the defect patterns of defect clusters.

Reliability is a measure that is closely related to the profitability of semiconductor fabrication along with yield. Conceptually, reliability can be decomposed into two failure modes: intrinsic and extrinsic failure mode [5]. Intrinsic failure mode is the inherited failure mechanism of the products, while extrinsic failure mode results from extraneous factors such as defects, bad assembly, misuse or environmental shock. Extrinsic failure mode shares the same source of failure as yield, the defects. Defects that don't cause an integrated circuit to fail until the end of the fabrication process or in the yield test may cause failure in the operation as the severity of defects increases, especially in the early stages of IC life. Thus, yield information can be utilized in extrinsic reliability estimation using a statistical defect growth model. It is known that well-designed and well-manufactured integrated circuits have a long useful life due to their high intrinsic reliability up to the mission time. Hence, extrinsic reliability has a significant influence on reliability and, consequently, on burn-in decisions that are made based upon information about early failures.

In this study, the spatial pattern of defects is modeled by the spatial point process and utilized in the yield modeling, defect pattern recognition for root-cause analysis and extrinsic reliability estimation from defect data. These methods will provide an unified framework of defect control and modeling in association with yield, yield learning and reliability.

The research objectives of this study are to (1) develop new yield and reliability models that directly attack defect density variation by location on the wafer and between wafers for better yield estimation and prediction and for fabrication process control applications, (2) develop an automated wafer defect data classifica-

tion algorithm that facilitates wafer map analysis especially for root-cause analysis in the yield learning process, and (3) develop an extrinsic reliability model based on yield information with new yield models and a statistical defect growth model.

The remainder of this dissertation is composed as follows. Chapter II reviews yield and the yield learning process. Several definitions of yield are presented and the basic elements of yield modeling such as defect/fault, defect density, defect density distribution, defect size distribution, fault ratio, critical area are discussed. Existing yield models are introduced and the limitations of these models are discussed. The notion of yield learning and the impact of the yield learning rate to the profitability of a semiconductor company is discussed.

Chapter III discusses basic notions of reliability. a definition of reliability and the mathematical properties of the reliability function are introduced. The aging property of a system is investigated using the failure rate function. Several measures related to reliability are defined. The notion of intrinsic and extrinsic reliability is presented. And existing yield-reliability relation models, which predict extrinsic reliability from yield data, are reviewed.

In chapter IV, yield models of integrated circuits using spatial point process are studied. The defect density variation by location within the wafer is modeled by the spatial nonhomogeneous Poisson process, and the defect density variation between wafers is modeled by the doubly stochastic point process. The assumption of point defect is generalized by the spatial marked point process in the general defect yield model. Construction of the intensity functions of spatial point processes is discussed, and estimation methods are presented. The results of a simulation are provided.

Chapter V studies a defect data analysis method of root-cause analysis in semiconductor fabrication. This model-based clustering method detects the existence

and the location of defect clusters generated from assignable causes, and it provides the characteristics of the clusters to facilitate wafer map analysis and eventually capture information about the process variations. In order to incorporate the various defect patterns on wafers, the bivariate normal distribution and the principal curve are applied.

Chapter VI studies the extrinsic reliability estimation model based on yield information. The extrinsic reliability model is based on the yield model studied in Chapter IV and the statistical defect growth model for describing the different defect growth behaviors of various kinds of defects. Defect growth factor models are discussed, and the property of the reliability function in terms of the failure rate function is investigated.

Chapter VII briefly summarizes this study.

CHAPTER II

YIELD AND YIELD LEARNING

In this section, the basic concepts used in yield modeling are introduced, and current yield models of integrated circuit(IC) are reviewed. Also, the notion of yield learning is discussed. For more information about yield models, refer to [6, 7, 8]. The ideas used in the existing yield models are useful for developing new yield models.

Yield can be estimated and predicted by yield models. Also, yield models can be used in the estimation of the manufacturing cost of products, process evaluation and investment decisions about new products. Yield models can compute the probability of at least one defect landing on an arbitrary die location and of any of those defects causing a failure of the IC. To compute this, we need the number of defects on a die area and the probability of a defect causing a circuit failure. The number of defects on a die area is governed by the defect density distribution. And the probability of a defect causing circuit failure, which depends on the interaction between the defect size and the IC topology, is evaluated using the defect size distribution and the critical area. The notation is defined as follows:

- D: defect density, a random variable
- $f(D)$: probability density function(PDF) of defect density D
- D_0, $D_0(x)$: average defect density and average defect density of size x
- X: defect size, a random variable
- $g(x)$: PDF of defect size X
- $A_c(x)$: critical area of defect size x

II.1. Yield

Yield is one of the most important measures of manufacturing processes [6]. It is closely related to the profitability of the manufacturer. The chip cost is defined as [5]

$$\text{chip cost} = \frac{\text{cost per wafer}}{\text{yield} \times \text{number of dies per wafer}}.$$

Thus, by increasing the yield, the semiconductor manufacturer can produce IC's at a lower production cost. The overall yield is defined as the product of the yield components:

$$Y_{\text{overall}} = Y_{\text{wafer}} Y_{\text{probe}} Y_{\text{assembly}} Y_{\text{final test}}.$$

The yield components are defined as

- $Y_{\text{wafer}} = \dfrac{\text{number of wafers completing parametric test}}{\text{number of wafers started}}$
- $Y_{\text{probe}} = \dfrac{\text{number of dies that pass the specification}}{\text{number of dies that reach electrical test}}$
- $Y_{\text{assembly}} = \dfrac{\text{number of chips assembled successfully}}{\text{number of chips that pass electrical test}}$
- $Y_{\text{final test}} = \dfrac{\text{number of chips that pass the final test}}{\text{number of assembled chips}}$

Among the yield components, Y_{wafer} and Y_{probe} are critical to the overall yield. This study is focused on modeling and improving Y_{probe}.

II.2. Defects and defect density

A defect is defined as a variation in quality that changes the IC topology. The main source of yield loss is defects, especially visible defects on the wafer. A visible defect is a visible object on the wafer such as particles, holes, etc. The causes of these visible defects are particles in the cleanroom, particles from equipment, or chemical

stains and scratches on the wafer caused by improper batch handling. The failure of a chip is caused by a defect on it. However, not all defects cause circuit failure. A fatal defect that causes circuit failure is called a fault. In order to compute the yield, information about the faults is needed, and it is obtained from defect data. Whether a defect becomes a fault is determined by its size and location on the die. The probability of a defect with an arbitrary size and location causing circuit failure is computed using the defect size distribution and the critical area, which are discussed in the next sections.

Denote $N(d\mathbf{s})$ to be the number of defects over an infinitesimal area around a point $\mathbf{s}$. The defect density is defined as [9]

$$D = \lim_{|d\mathbf{s}|\to 0} \frac{P\{N(d\mathbf{s}) = 1\}}{|d\mathbf{s}|}.$$

Here, the probability that more than one defect occurs in a very small area is assumed to be negligible. From the definition, the following are obtained:

$$D_0(x) = D_0 g(x),$$

and, consequently,

$$D_0 = \int_0^\infty D_0(x)dx,$$

II.3. Defect size distribution

The size of a defect plays an important role in determining whether the defect becomes a fault or not. Defect size is considered a random variable by the defect size distribution. And the critical area of an arbitrary defect size, the average critical area, is the integrated product of the critical area and the defect size distribution. For inference of the defect size distribution, a specially constructed test structure is

used since an optical measurement of the defect size is not reliable. It is assumed that there is a dominating defect size in the observation frequency, i.e., there is a peak in the probability density function at a particular defect size and the density diminishes on both sides of it [10, 11]. The probability density follows the power law for defects smaller than the optically resolvable size x_0 and the inverse power law for defects bigger than x_0 [10, 12]:

$$g(x) = \begin{cases} \dfrac{(q+1)(p-1)}{q+p} \dfrac{x^q}{x_0^{q+1}}, & \text{if } x \le x_0, \\ \dfrac{(q+1)(p-1)}{q+p} \dfrac{x_0^{p-1}}{x^p}, & \text{if } x > x_0, \end{cases}$$

where $p \neq 1$, $q > 0$. From experiments at IBM, Stapper [11] finds that the probability density of defects smaller than x_0 increases linearly, i.e., $q = 1$:

$$g(x) = \begin{cases} \dfrac{2(p-1)x}{(p+1)x_0^2}, & \text{if } x \le x_0, \\ \dfrac{2(p-1)x_0^{p-1}}{(p+1)x^p}, & \text{if } x > x_0. \end{cases}$$

The distribution is sketched in Figure 1.

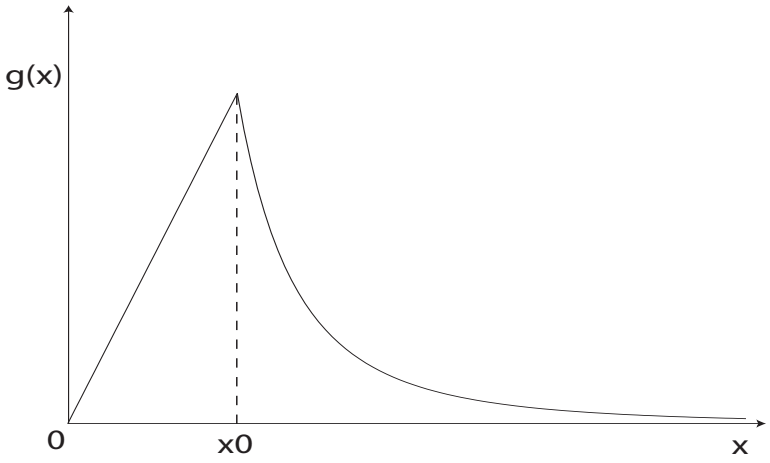


Figure 1: Defect size distribution.

II.4. Critical area

Compound Poisson yield models require the number of faults on a die. However, what we gain from wafer inspection is information about the defects. We need to compute the probability of a defect with an arbitrary size and location causing a circuit failure. This probability can be estimated by the fault probability defined as

$$\Phi \equiv \frac{\text{number of faults out of defects}}{\text{number of defects}}.$$

Whether a defect becomes a fault depends on the location and the size of the defect associated with the circuit design. Figure 2 illustrates the different sensitivities of two chip designs. Wider conductive lines and wider pitches lead to lower fault-to-defect ratios. Given a defect with particular size and location on a die, we can probably find out whether or not it will cause a failure. However, since it is impossible to examine all of the defects, a probabilistic method using the notion of critical area is considered.

Critical area is defined as the collection of points in which the existence of the center of a defect causes the failure of the IC [11]. The critical area is a function of defect size x, and it is denoted as $A_c(x)$. The following are illustrations of the critical area with simple circuit design. Consider the missing material defects on a circuit with a conducting line in the middle of the circuit. Defects smaller than w, the width of the conducting line, don't cause the circuit to fail, and defects larger than $H + w$ cause circuit failure wherever they are located on the circuit. For defects sizes between w and $H + w$, the critical area is found in the mid part of the conducting line, and it depends on the defect size. We have the critical area of Figure 3:

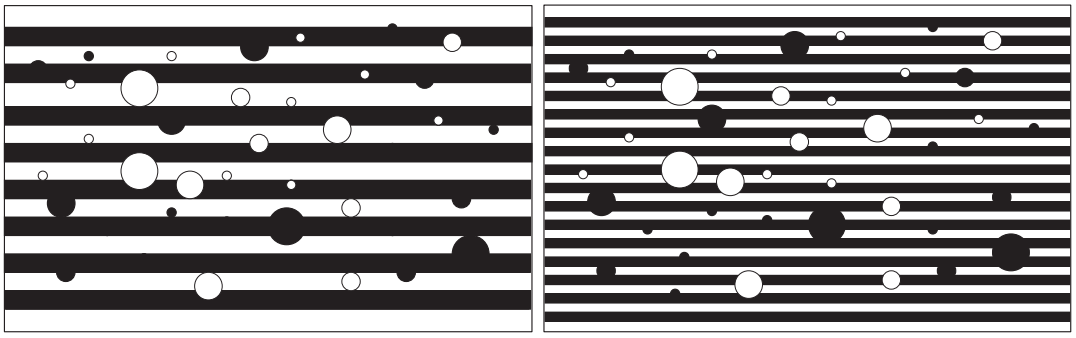


Figure 2: Defects and faults.

Two different designs of chips. Both chips have the same defects on them. 20 black dots represent extra material causing shorts and 20 circles represent missing material causing open. In the left figure, 2 of the extra defects and none of the missing defects become faults. On the other hand, in the right figure, 9 of the extra defects and 9 of missing defects become faults.

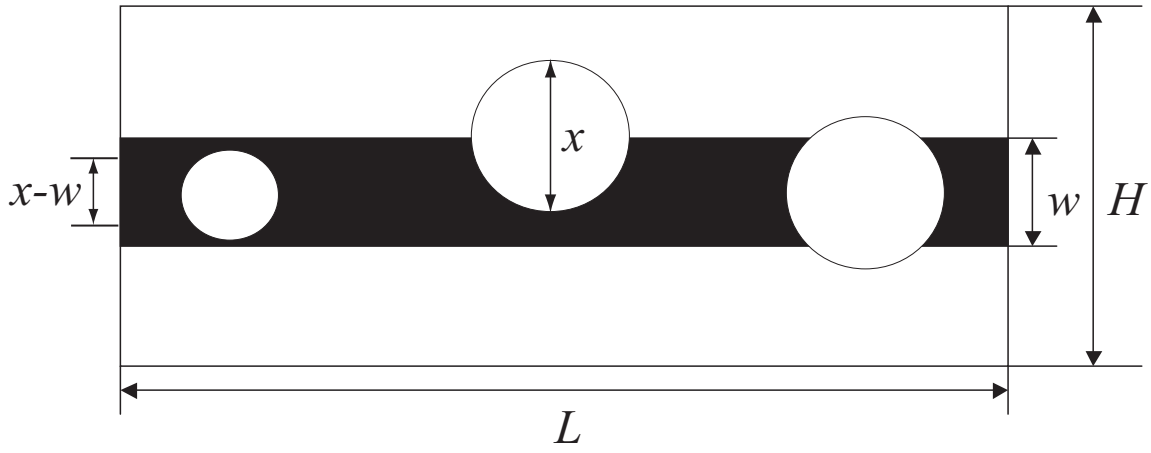


Figure 3: Chip with single conduction line.

The left two defects don't cause failure of the chip: the first one is smaller than the critical dimensions of the chip, while the second that is bigger than the width of the conductive line is located out of the critical area. The third defect causes a circuit failure. The critical area is found along the center of the conductive line.

$$A_c(x) = \begin{cases} 0, & \text{for } 0 \le x < w, \\ L(x-w), & \text{for } w \le x < H + w, \\ LH, & \text{for } x \ge H + w. \end{cases}$$

This kind of analytical model of the critical area cannot be applied to actual ICs since the geometry is too complicated, and the shape of defects is not perfectly round. Instead, simulation methods are used in practice [13].

Now, the fault probability of defect size x can be defined using the critical area [12, 11]:

$$\Phi(x) \equiv \frac{A_c(x)}{A},$$

where A is the die area. And the number of defects can be computed

$$AD(x)dx.$$

Thus, the number of faults is obtained by the product of the number of defect and faults probability:

$$\Phi(x)AD(x)dx = A_c(x)D(x)dx.$$

Define the average critical area as

$$A_c \equiv \int_o^\infty A_c(x)g(x)dx,$$

then the average number of faults of all of sizes on a die is

$$\mu = \int_0^\infty A_c(x)D_0(x)dx = D_0 \int_0^\infty A_c(x)g(x)dx = A_cD_0. \tag{2.1}$$

II.5. Yield models

The Poisson model which is derived from the Poisson distribution is one of the simplest models. Given the average number of faults on a die, $\mu = A_c D_0$, the probability that a die contains n faults can be computed from the Poisson distribution:

$$P_n = \frac{e^{-\mu}\mu^n}{n!},\ n = 0, 1, \cdots .$$

The Poisson yield model is defined as

$$Y \equiv P_0 = e^{-\mu}. \tag{2.2}$$

The binomial model is derived from the binomial distribution:

$$P(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}.$$

$P(k)$ is the probability of k success in n trials with p probability of success. In the binomial yield models, n represents the total number of defects on the wafer computed $A_w D_0$, where A_w is the area of wafer. And the probability of success for each trial p is computed by A_c/A_w. The yield model which is defined as $P(0)$ with n and p obtained as above is

$$Y \equiv \left(1 - \frac{A_c}{A_w}\right)^{A_w D_0}.$$

This model converges to the Poisson yield model, equation (2.2), as $A_w \to \infty$.

Poisson and binomial yield models assume that defects on a die are uniformly distributed (however, it is worth noting that it doesn't mean that the distances between them are all the same). These models provide relatively accurate yields for small die sizes of less than 0.25 cm^2 [6]. But as the chip size increases, they

underestimate the yield because of the assumption of the uniform distribution of the defects.

The defects on IC products are known to cluster, and this causes variation in defect density. In order to take the defect density variation into account, compound Poisson yield models are introduced. In these yield models, the defect density is considered as a random variable, and yield is defined as the expectation of equation (2.2) against the defect density distribution. The yield is then

$$Y = \int_0^\infty e^{-A_c D} f(D) dD. \tag{2.3}$$

Different compound Poisson yield models are variations with different defect density distributions $f(D)$.

It can be assumed that the defect density distribution has a peak and diminishes symmetrically on both sides. This distribution can be modeled by the normal distribution. However, a compound yield model in a closed form with the normal distribution does not exist. If the triangular distribution, which is the approximation of the normal distribution, is compounded to the Poisson yield model:

$$f(D) = \begin{cases} \frac{D}{D_0}, & \text{for } 0 \le D < D_0, \\ \frac{2D_0 - D}{D_0^2}, & \text{for } D_0 \le D \le 2D_0, \end{cases}$$

then the resulting Murphy's [14] yield model is

$$Y = \left(\frac{1 - e^{-A_c D_0}}{A_C D_0} \right)^2.$$

Seeds [15] proposes a yield model with the exponential distribution:

$$f(D) = \frac{e^{-D/D_0}}{D_0}, \ D \ge 0.$$

This distribution assumes that the chance of high defect density decreases rapidly. The resulting yield model is

$$Y = \frac{1}{1 + A_c D_0}.$$

The advantage of this model is that the shape of the distribution can be adjusted according to the mean and variance of the defect density. This is the first work that points out the statistical aspect of defect density variation [6].

If we assume the uniform distribution on $[0, 2D_0]$, then we have

$$Y = \frac{1 - e^{-2A_c D_0}}{2A_c D_0}.$$

This model assumes that the defect density is uniform up to $2D_0$. It can be considered as an approximation of the normal distribution [16].

Stapper [17] proposes a yield model with the half Gaussian distribution defined as

$$f(D) = \frac{2}{\pi D_0} \exp\left\{-\frac{1}{\pi}\left(\frac{D}{D_0}\right)^2\right\},\ D \geq 0,$$

and the resulting yield model is written as

$$\begin{aligned} Y &= \frac{2}{\sqrt{\pi}} \exp\left\{\frac{\pi (A_c D_0)^2}{4}\right\} \int_{\sqrt{\pi} A_c D_0/2}^{\infty} e^{-u^2} du \\ &= \exp\left\{\frac{\pi (A_c D_0)^2}{4}\right\} \operatorname{erfc}\left(\frac{\sqrt{\pi} A_c D_0}{2}\right), \end{aligned}$$

where the complementary error function is defined as

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^{\infty} e^{-u^2} du.$$

In some applications, the actual yield is observed between Seed's and Murphy's yields and can be calculated from the averaged yields of those two yield models. In those cases, this yield model represents the averaged yield [17].

One of the widely used compound yield models is the negative binomial model which uses the gamma distribution defined as

$$f(D) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} D^{\alpha-1} e^{-D/\beta}, \ D \geq 0.$$

The yield model is

$$\begin{aligned} Y &= \int_0^{\infty} e^{-A_c D} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} D^{\alpha-1} e^{-D/\beta} dD \\ &= (1 + A_c\beta)^{-\alpha} \\ &= \left(1 + \frac{A_c D_0}{\alpha}\right)^{-\alpha}. \end{aligned} \tag{2.4}$$

The parameter, α, is called the clustering factor. By varying the value of α, the negative binomial yield model covers all of the range of defect clustering effects. The smaller the α, the higher the degree of clustering. With different α values, the negative binomial model can represent several yield models [6, 8]. If $\alpha \to \infty$, then equation (2.4) becomes equal to the Poisson yield model. The range of α in practice is known to be from 0.3 to 5 [7]. Figure 4 sketches the defect density distributions used in yield modeling and the yield models derived from those distributions are summarized in Table 1. Figure 5 shows plots of yield vs average number of faults per chip.

An approach other than the compound Poisson yield models was studied by Berglund[18]. Berglund points out that the yield loss which occurs on multiple die locations can be explained by defects that have comparable, or even larger, defects than the die size and thus destroy several overlapped dies. He combines the second equation of equation (2.1) and equation (2.2) to get

$$Y = \exp\left\{-\int_0^{\infty} A_c(x) D_0(x) dx\right\}, \tag{2.5}$$

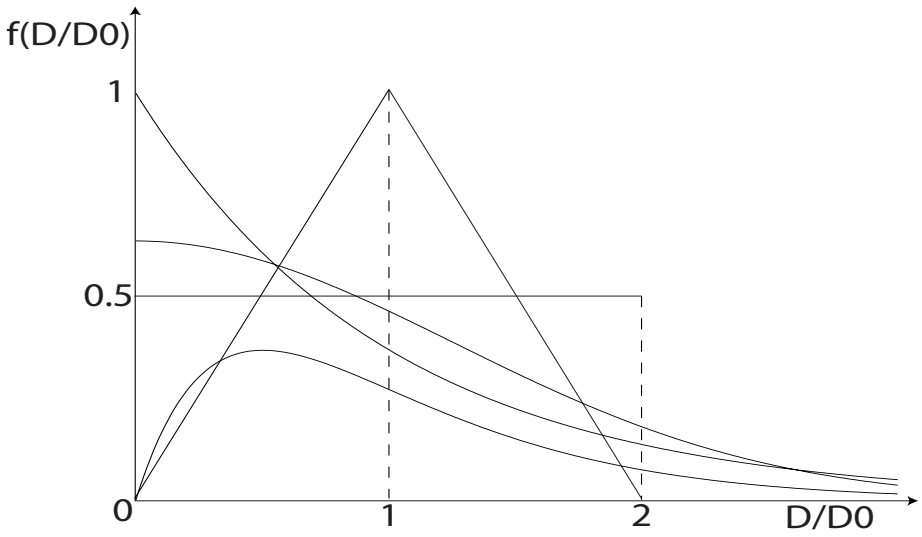


Figure 4: Defect density distributions: half Gaussian, exponential, triangular, uniform and gamma distribution($\alpha = 2$).

with

$$A_c(x) = LW + (L + W)x + \pi x^2/4,$$

where L and W are the width and length of the die, respectively. And equation(2.5) becomes

$$Y = Y_d Y_s = \exp\{-LWD_0\} \exp\{-(L + W)E[X]D_0 - E[X^2]D_0\}.$$

Y_d is equivalent to the Poisson yield model, and Y_s can be interpreted as the yield loss caused by large-size defects.

II.6. Limitations of existing yield models

Defect density variation is modeled by the defect density distribution in compound yield models. It describes the mean and variance of defect density. However, it doesn't provide any information about defect density variation by location on

Table 1: Compound Poisson yield models

Name	$f(D)$	Y
Seed's	$\frac{e^{-D/D_0}}{D_0},\ D \geq 0$	$\frac{1}{1+A_cD_0}$
Murphy's	$\begin{cases} \frac{D}{D_0}, & \text{for } 0 \leq D < D_0, \\ \frac{2D_0-D}{D_0^2}, & \text{for } D_0 \leq D \leq 2D_0 \end{cases}$	$\left(\frac{1-e^{-A_cD_0}}{A_CD_0}\right)^2$
Half-Gaussian	$\frac{2}{\pi D_0}\exp\left\{-\frac{1}{\pi}\left(\frac{D}{D_0}\right)^2\right\},\ D \geq 0$	$\exp\left\{\frac{\pi(A_cD_0)^2}{4}\right\}\text{erfc}\left(\frac{\sqrt{\pi}A_cD_0}{2}\right)$
Uniform	$\frac{1}{2D_0},\ 0 \leq D \leq 2D_0$	$\frac{1-e^{-2A_cD_0}}{2A_cD_0}$
Negative binomial	$\frac{1}{\Gamma(\alpha)\beta^\alpha}D^{\alpha-1}e^{-D/\beta},\ D \geq 0$	$\left(1+\frac{A_cD_0}{\alpha}\right)^{-\alpha}$

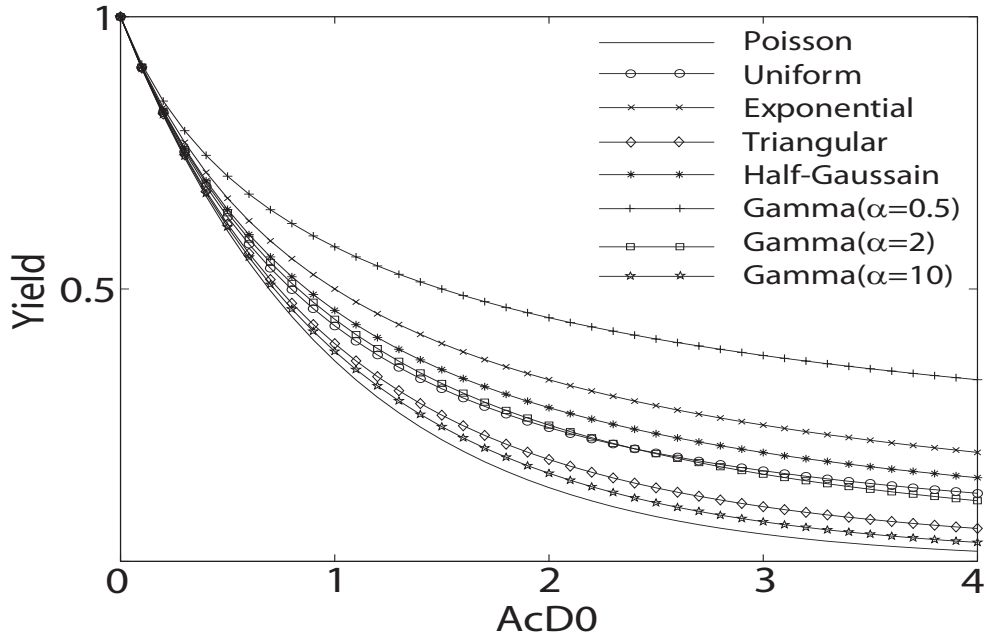


Figure 5: Yield models.

the wafer [19]. Also, defect density distribution is hard to infer from the actual wafers [13]. Due to these limitations, the defect density distributions estimated from different layers of the ICs, different process steps or different defect types cannot be incorporated to get an overall defect density distribution. And, since the yield models fail to describe the defect pattern, the yield itself and the parameters of these yield models are not suitable measures for process control [20].

II.7. Defect monitoring and yield learning

Yield learning is a procedure of removing the sources of defects to achieve a desired yield for the process. In the dynamic semiconductor industry, high yield alone doesn't guarantee the success of a company. The target yield must be achieved on time for increased profit to result. The pressure of rapid yield learning comes from increased competition, the shortened life cycles of IC products and rapid advances in technology. As Weber [3] points out, several months difference in a yield learning schedule can make a difference of several billion dollars in net profit. Finding the sources of yield loss, called root-cause analysis, calls for the statistical process control methods along with summarized measures of the process parameters and measurements of the products. Even with abundant data, however, root-cause analysis is still in an art form. Huge amounts of data and unclear relationships between the yield loss mechanisms and their measurements are the main obstacles inhibiting root-cause analysis in the semiconductor industry. The semiconductor manufacturing process comprises several hundred processes including epitaxy, deposition, diffusion, etching, etc, and each process has more than twenty parameters. Thus, finding the relation between these parameters and yield is very challenging [2].

Defects are the result of process variations and are the main source of yield

loss. Consequently, they contain useful information about process variations for yield learning. Information about defects is obtained from in-line and off-line inspection. Normally, defects are categorized according to their sources. In this study, the sources of variation are categorized into two groups, the random causes and the assignable causes, for the purpose of process control. The random causes produce defects on ICs by chance. The defects generated from random causes are called global defects. These defects are observed all over the wafer. It is known that a defect is more likely to occur on the edge of the wafer due to thermal variation in the annealing process and mechanical stress. The assignable causes, on the other hand, generate defects systematically, and, thereby, generate sets of aggregated defects, or clusters. These defects are called local defects.

II.8. Summary

In this chapter, basic notions of yield and yield learning are introduced, and existing yield models are reviewed. Variations in semiconductor manufacturing processes still keep the yield modeling from being precise in spite of numerous attempts to improve them during the last several decades. Defect density variation is modeled by the compound distribution which describes the randomness of the defect density. Different defect density distributions lead to different yield models. And those yield models have different properties and represent different defect clustering effects. For this reason, the negative binomial yield model is powerful since it can represent the full range of defect clusterings. However, defect density distribution doesn't provide spatial information about defect density variation, and the lack of spatial information about defect patterns in the yield models restricts them from being fully utilized in the yield learning process, especially in root-cause analysis which is conducted based

on defect information.

CHAPTER III

RELIABILITY

Reliability is a factor that is related to customer satisfaction and, consequently, to a manufacturer's profit. Reliability is defined as the probability that a system works properly up to a specified time under specified working conditions [5], and the reliability function is a function of the mission time t. If the time is assumed to be fixed at specific value, the reliability is called static reliability. Since the actual failure mechanism can never be fully analyzed, statistical inference is widely used for the reliability estimation of a system. The following are the basic concepts of reliability derived from the life time distribution of a system.

III.1. Reliability concepts

Denote a random variable, T, to be the life time of a system and F to be the cumulative distribution function of T:

$$F(t) = \Pr\{T \leq t\}.$$

Then, the reliability function $R(t)$ is

$$\begin{aligned} R(t) &= \Pr\{\text{system works without failure up to time } t\} \\ &= \Pr\{T > t\} \\ &= 1 - \Pr\{T \leq t\} \\ &= 1 - F(t). \end{aligned}$$

Define the conditional reliability of age t as

$$R(x|t) \equiv \Pr\{T > t + x | T > t\} = \frac{R(t+x)}{R(t)}, \text{ if } R(t) > 0.$$

The conditional probability of failure during the next interval x of age t is

$$F(x|t) \equiv \Pr\{t < T \leq t + x | T > t\} = \frac{F(t+x) - F(t)}{R(t)} = 1 - R(x|t).$$

The failure rate(or hazard rate) function is defined as

$$h(t) \equiv \lim_{x \to 0} \frac{1}{x} \frac{F(t+x) - F(t)}{R(t)} = \lim_{x \to 0} \frac{1}{x} [1 - R(x|t)], \text{ if } R(t) > 0.$$

If $f(t) \equiv \dfrac{dF(t)}{dt}$ exists,

$$h(t) = \frac{f(t)}{R(t)}. \tag{3.1}$$

From equation (3.1) and $f(t) = -\dfrac{dR(t)}{dt}$,

$$\int_0^t h(x)dx = -\log R(t).$$

Thus,

$$R(t) = \exp\left\{ -\int_0^t h(u)du \right\}.$$

The failure rate function represents the aging characteristic of a system. If a system ages, i.e., the system deteriorates in time, the conditional reliability of age t,

$$R(x|t) \text{ is decreasing in } t, \ \forall x \geq 0, \tag{3.2}$$

then, we have

$$h(t) = \lim_{x \to 0} \frac{1}{x} [1 - R(x|t)] \text{ is increasing in } t.$$

Conversely, if the failure rate is increasing in t, then

$$R(x|t) = \exp\left\{-\int_t^{t+x} h(u)du\right\} \text{ is decreasing in } t,\ \forall x \geq 0.$$

Thus, if a failure rate function exists, increasing the conditional reliability is equivalent to increasing the failure rate function.

A distribution with decreasing conditional reliability is defined as an increasing failure rate(IFR) distribution, and a distribution with increasing conditional reliability is called a decreasing failure rate(DFR) distribution [21].

The mean residual life of age t is defined as

$$m(t) \equiv E[T - t|T > t] = \frac{\int_t^{\infty} R(u)du}{R(t)},\ \text{if } R(t) > 0.$$

An increasing failure rate function is a stronger condition than an increasing mean residual life [22].

Electronic devices including ICs are known to have the bathtub-shaped failure rate curve. The bathtub-shaped failure rate curve can be broken down into two independent failure rate curves with competing risk reliability model. One failure rate curve results from the extrinsic failure mode which has an IFR failure distribution. And the other DFR curve represents the intrinsic failure mode. Figure 6 shows a typical failure rate curve of IC products.

III.2. Burn-in

Burn-in is a technique for improving the reliability of a product by screening out weak products before they reach customers. This effect is achieved by applying intentional aging of the products with stress. Burn-in has long been used as a method for reducing early failures. However, burn-in has several side effects: reducing the

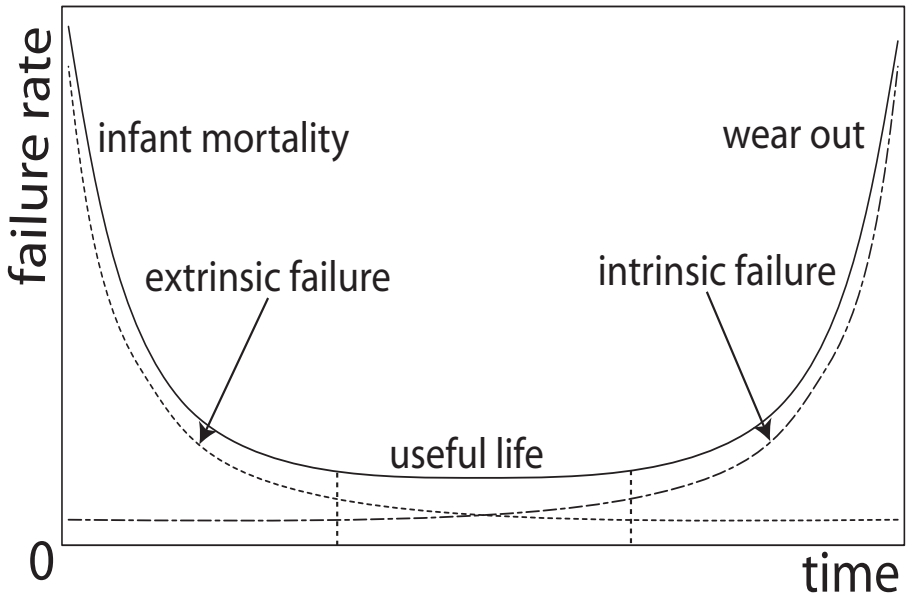


Figure 6: Bathtub-shaped failure rate curve.

mean residual life of the system and associated costs. If products are well designed and manufactured, conditional reliability is not increasing in burn-in time, it merely reduces the mean residual life. Thus, conditions for the necessity of burn-in should be examined carefully.

Conditional reliability and the mean residual life are used for criteria of burn-in [23]. If we use either conditional reliability or mean residual life for a burn-in criterion, increasing the failure rate is a sufficient condition for the necessity of burn-in. However, the optimal burn-in times for two criteria are not necessarily the same. Thus the level, duration and stress of burn-in should be wisely determined to maximize the burn-in effect.

Mean residual life is a long term measure compared to conditional reliability. The mission time of a system should be considered. Even with an increased mean residual life burn-in, if the conditional reliability is up to mission time $t_b + x$, the procedure has little impact on profit.

III.3. Yield-reliability relation models

Reliability estimation from yield data has been studied by several researchers [24, 25, 26, 27]. Yield and reliability are known to have positive correlation [28]. However, the exact modeling of the relationship between yield and reliability is far from precise as yet. Before I proceed to the relation models, let's examine the theoretical aspects.

Reliability is a function of time. A system is perfect at the beginning of the operation, i.e., $R(0) = 1$. However, the point at $t = 0$ depends on the epoch of the end of the production and the duration of the tests and burn-in. To avoid the difficulty of choosing $t = 0$, let's define a random variable S_t to represent the state of a product at epoch t:

$$S_t = \begin{cases} 1, \text{ if product is working at } t, \\ 0, \text{ if product is failed at } t. \end{cases}$$

and a probability function Q such that

$$Q(t) = \Pr\{S_t = 1\}.$$

Q is a nonincreasing function in t.

Let t_m denote the time just after production. Then, yield can be written as

$$Y = Q(t_m).$$

If we redefine the reliability function with conditional probability, then

$$R(t) = Q(t_m + t | t_m) = \frac{Q(t_m + t)}{Y}, \text{ if } Y > 0.$$

Unfortunately, the general function of Q doesn't exist since the failure mechanisms of

yield and reliability are not necessarily the same. However, if we consider the extrinsic reliability which is determined by the defects introduced during the manufacturing processes, we may be able to find Q.

The following is review of existing yield-reliability relation models. Existing models can be classified into two approaches: the latent defect approach and the defect growth approach. In the reliability defect(or latent defect) approach, it is assumed that there exist a certain portion of reliability defects that do not damage ICs during manufacturing, but cause the failure of ICs in operation. Denote D_r and D_y to be the reliability defect density and the yield defect density, respectively. And we assume the relation:

$$D_r = \alpha D_y.$$

Then, the reliability is obtained using the yield model [24]:

$$R = e^{-A_c D_r} = e^{-A_c \gamma D_y} = Y^{\gamma}. \tag{3.3}$$

By introducing clustering and the edge excursion factor $M > 0.9$, the reliability is [29, 28]

$$Y = \left(\frac{Y}{M}\right)^{\gamma}.$$

A similar, but essentially the same, approach uses the reliability critical area which has the relation [25]:

$$A_r = \alpha A_c.$$

When a Poisson yield model is used, this model provides the same result as equation (3.3) if $\alpha = \gamma$. Since D_r is not a function of time, the reliability in these models is static.

The defect growth model is applied to gate oxide [26, 27]. The defect growth

rate is derived from the defect-related time-to-breakdown of oxide:

$$t_{BD} = \tau \exp\left\{\frac{G}{V}(w - x)\right\},$$

where t_{BD}: the breakdown time in seconds, w: the oxide thickness, τ: a constant($\approx 10^{-11}$ sec), V: the voltage across the oxide and G: a constant($\approx$ 350 MV/cm). T. Kim et al. [27] derive the reliability critical area of defect size x, $A_r(x)$ using the threshold defect growth, w_g, at a given mission time, t:

$$w_g = \frac{V}{G} \ln\left(\frac{t}{\tau}\right).$$

The reliability critical area for a single conductive line with length L is

$$A_r(x) = \begin{cases} 0, & 0 \le x \le w - w_g, \\ L(x + 2w_g - w), & w - w_g \le x \le w, \\ 2Lw_g, & w \le x. \end{cases}$$

K.O. Kim et al. [26] assume that there exists a certain threshold defect size x^* and that defects larger than x^* cause the gate oxide which is to fail at the end-of-line test. For an arbitrary defect on the gate oxide, given a mission time t, the probability that the defect won't break down the oxide, $\Pr\{T \le t\}$ is computed by

$$\Pr\left\{X \ge w - \frac{V}{G} \ln \frac{t}{\tau}\right\}$$

from the defect size distribution, $g(x)$. Thus, the average number of defects that pass the end-of-line test but fail the oxide in operation before t is

$$AD \Pr\left\{T \le t, T \ge \tau \exp\left\{\frac{G}{V}(w - x^*)\right\}\right\}, \tag{3.4}$$

where A is the gate oxide area and D is the defect density. The reliability is obtained from Poisson or negative binomial yield model with (3.4).

III.4. Summary

The aging characteristic is described by the failure rate function of the life distribution. According to a property of the failure rate function, a life distribution can be classified into a decreasing failure rate or an increasing failure rate distribution. The typical bathtub-shaped failure rate function of ICs can be modeled by summation of two independent monotone failure rate functions based on the competing risk model. Extrinsic reliability, which has a decreasing failure rate, is determined by the defects. Since yield and extrinsic reliability have the same source of loss, extrinsic reliability can be estimated from defect data or yield.

CHAPTER IV

YIELD MODELING OF INTEGRATED CIRCUITS BASED ON SPATIAL NONHOMOGENEOUS POISSON PROCESS

Yield models of ICs using spatial point process are studied. The point defect yield model assumes that regardless of the defect size, every defect causes a circuit failure . The defect density variation by location is modeled by spatial nonhomogeneous Poisson process, and defect density variation by wafer is modeled using a doubly stochastic point process. The assumption of the point defect is generalized by spatial marked point process in the general defect yield model. Construction of the intensity functions of the spatial point processes are discussed and estimation methods are presented. Validation of the new yield models with simulated data is also provided.

IV.1. Introduction

Yield, the measure of the success of a manufacturing process, is a particularly important characteristic in the semiconductor industry because the yield of IC products is closely related to profitability [30]. Yield is estimated and predicted by yield models. With these models, variations in the manufacturing process that cause yield loss can be estimated. Also, yield models are useful for estimating the manufacturing cost of products under development, for facilitating IC design by providing yield predictions at the design stage and for improving manufacturing processes by providing a measure of the success of a process [6]. Although there have been many attempts to model the yield of ICs since the 1960s, accurate yield modeling remains elusive.

The most widely used yield models are the compound the Poisson yield models that compound Poisson distribution with other distributions. Refer to [6, 7, 8].

These yield models have some limitations such as (1) the defect density distribution and defect size distribution are hard to infer from the actual wafers [13] or (2) they can't provide information about defect density by location [19]. Due to these limitations, the defect densities estimated by each processes cannot be incorporated to find the overall defect density. Also, since these yield models fail to describe the defect pattern, the yield itself, and the parameters of these yield models are not suitable measures for process control [20]. Hence a new technique for yield modeling and statistical process control that directly deals with spatial defect density variation is needed.

Modeling the spatial distribution of events has been a concern in many fields [31, 32]. Spatial point processes have been widely studied in epidemiology, meteorology, astronomy and so on. The spatial Poisson process is a point process which describes spatial point patterns that are generated by unknown random mechanisms [32]. The defects on a wafer can be considered as a realization of a spatial nonhomogeneous Poisson process(NHPP).

In this dissertation, I propose new yield models of ICs based on the spatial NHPP. In the point defect yield model, it is assumed that every defect on a die causes a circuit failure. Thus, the spatial defect pattern is the main interest. The point defect yield model can be generalized with the spatial marked point process to incorporate various sizes of defects. In the general defect yield model, the size of the defect is considered the marked value. The spatial defect pattern is modeled by an intensity function of the process defined as equation (6.4). The new yield models have several advantages as follows: (1) a huge amount of wafer defect data can be reduced by the coefficients of the intensity function, (2) by the additivity of the intensity function, the intensities of the different defect types and those estimated for different processes can be added to describe the spatial defect distribution of the

whole process and (3) they provide more sensitive measures of the variations of the manufacturing processes than does yield alone since the intensity function contains information about the average defect number and defect clustering.

The defect pattern on a wafer can be considered a superposition of defect patterns from several independent defect generation mechanisms. Modeling this variation with a single intensity function is not appropriate since it requires numerous parameters. To overcome this limitation, I consider clustering methods for constructing and estimating an intensity function that is accurate and flexible.

Since defect data in the semiconductor industry is proprietary, the performance of the new yield models are evaluated by simulation.

The remainder of this chapter is arranged as follows: In section IV.2, existing yield models are reviewed. The basic concepts in compound Poisson yield models in particular are examined. Section IV.3 presents the background of the spatial point process. Section VI.3 discusses the point defect yield model using NHPP and general defect yield model that deals with various sizes of defects. Intensity function estimation and modeling using clustering is examined in section IV.5. Simulation methods and results are presented in section IV.6. And section VI.7 concludes the chapter.

IV.2. Yield model review

In this section, current yield models are reviewed. For more information on yield models, see [6, 7, 8]. The concepts used to develop these yield models are also relevant to the new yield models. Yield models compute the probability of a event that at least one defect lands on an arbitrary die location and that it causes a failure of the IC. A defect is defined as a variation in quality that changes the IC topology.

The failure of a chip is caused by a defect on it. However, not all the defects cause failures. A defect which causes the failure of an IC is called a fault. If the size of the defect on a die is given, then it is the location that differentiates a defect from a fault. A collection of locations where the existence of the center of a defect causes a failure is called a critical area. Whether a defect on a die causes a failure of an IC depends on the factors that affect the size of critical area, e.g., defect size, defect location and the layout of IC such as the gate oxide thickness, the length and width of the conducting path and the thickness of the insulation layers, etc. The notation is as follows:

- D: defect density, a random variable
- D_0, $D_0(x)$: average defect density and average defect density of size x
- $g(x)$: probability density function(PDF) of defect size X
- $f(D)$: PDF of defect density
- μ: average number of faults
- $A_c(x)$: critical area of defect size x

Let's begin with the definition of defect density. Denote $N(d\mathbf{s})$ to be the number of defects over the infinitesimal area around a point $\mathbf{s}$. Defect density is defined as [9]:

$$D = \lim_{|d\mathbf{s}|\to 0} \frac{P\{N(d\mathbf{s}) = 1\}}{|d\mathbf{s}|}.$$

Here the probability that more than one defect occurs at the same location is assumed to be negligible. From the definitions, the following relations are obtained:

$$D_0(x) = D_0 g(x).$$

So,

$$D_0 = \int_0^\infty D_0(x)dx,$$

The average critical area is obtained as

$$A_c \equiv \int_o^\infty A_c(x)g(x)dx. \tag{4.1}$$

Then the average number of faults on a die is

$$\mu = \int_0^\infty A_c(x)D_0(x)dx = D_0 \int_0^\infty A_c(x)g(x)dx = A_c D_0. \tag{4.2}$$

The Poisson model which is derived from the Poisson distribution is one of the simplest models. This model assumes that defects on a die are distributed uniformly and randomly. Given $\mu = A_c D_0$, the probability that a die contains n faults can be computed from the Poisson distribution:

$$P_n = \frac{e^{-\mu}\mu^n}{n!},\ n = 0, 1, \cdots . \tag{4.3}$$

Thus the Poisson yield model is

$$Y \equiv P_0 = e^{-\mu}. \tag{4.4}$$

This model provides relatively accurate yields for small die sizes [6]. But as the chip size increases, it underestimates the yield because of the assumption of random defects. Defects in IC products are known to cluster, which causes variations in defect density.

In order to take defect density variation into account, compound Poisson yield models are introduced. In these yield models, defect density is considered as a random variable and yield is defined as the expectation of equation (4.4) against the

defect density distribution. The yield is then

$$Y = \int_0^\infty e^{-A_c D} f(D) dD. \tag{4.5}$$

Different compound yield models are variations with different defect density distributions $f(D)$. One of the widely used compound yield models is the negative binomial model which uses the gamma(α, β) defect density distribution:

$$f(D) = \frac{1}{\Gamma(\alpha)\beta^\alpha} D^{\alpha-1} e^{-D/\beta}, \ D \geq 0.$$

Then we have

$$\begin{aligned} Y &= \int_0^\infty e^{-A_c D} \frac{1}{\Gamma(\alpha)\beta^\alpha} D^{\alpha-1} e^{-D/\beta} dD \\ &= (1 + A_c\beta)^{-\alpha} \\ &= \left(1 + \frac{A_c D_0}{\alpha}\right)^{-\alpha}. \end{aligned} \tag{4.6}$$

One of the parameters of the gamma distribution, α, is called the clustering factor. By varying the value of α, the negative binomial yield model covers all of the range of defect clustering. The smaller the α, the higher the degree of clustering. The range of α in practice is from 0.3 to 5 [7]. With different α values, the negative binomial model can represent several yield models [6, 8]. If $\alpha \to \infty$, then equation (4.6) becomes equal to the Poisson yield model.

An approach that is different from the compound Poisson yield models was studied by Berglund[18]. Berglund points out that yield loss that occurs on multiple die locations can be explained by defects that have comparable, or even larger, defects than the die size, thus they destroy several overlapped dies. He combines the second

equation of equation (4.2) and equation (4.4) to get

$$Y = \exp\left\{-\int_0^\infty A_c(x)D_0(x)dx\right\}, \tag{4.7}$$

with

$$A_c(x) = LW + (L+W)x + \pi x^2/4,$$

where L and W are the width and length of a die, respectively. And equation(4.7) becomes

$$Y = Y_d Y_s = \exp\{-LWD_0\}\exp\{-(L+W)E[X]D_0 - E[X^2]D_0\}.$$

The first term of the last equation is equivalent to the Poisson yield model, and the second term can be interpreted as the yield loss caused by the large-size defects.

As shown in Berglund's yield model which is constructed without a defect clustering effect, the clustering effect of defective chips is explained by an effect that affects more than one chip area. Likewise, but from a slightly different perspective, the clustering of defects on a wafer can be explained by a large defect density on a portion of the wafer area.

IV.3. Spatial nonhomogeneous Poisson process

In order to estimate yield, three things must be known: the spatial distribution of the defect density, the defect size distribution and the interaction of these two variables with the layout of the IC. These are the factors of compound Poisson yield models as described in the previous section. The defect density distribution, which is not well modeled in compound Poisson yield models, can be nicely described by spatial point processes. In the new yield models based on spatial NHPP, defect clustering is

explained by the higher defect density areas that produce aggregated defects. Useful concepts and definitions of the spatial point process are summarized below from Diggle [32] and Cressie [31].

The spatial process is used to model spatial data. The general model of spatial process is

$$\{Z(\mathbf{s}) : \mathbf{s} \in D\},$$

where $Z(\mathbf{s})$ is a random quantity at a location $\mathbf{s} \in D$. In many cases, among many possible measures of events, the locations of events $\mathbf{s}_1, \mathbf{s}_2, \ldots$ on $D \subset \mathcal{R}^d$ are the main concern. The locations can be modeled by the spatial point process as follows. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let Φ be a collection of locally finite counting measures on $D \subset \mathcal{R}^d$. On Φ, define $\mathcal{N}$, the smallest σ-algebra generated by sets of the form $\{\phi \in \Phi : \phi(B) = n\}$ $\forall B \subset D$ and $\forall n \in \{0, 1, 2, 3, \ldots\}$. Then a spatial point process on D is a measurable mapping of $(\Omega, \mathcal{F})$ into $(\Phi, \mathcal{N})$. If an event located at $\mathbf{s} \in \mathcal{R}^d$ is marked by $x \in \mathcal{A}$, then $(\mathbf{s}, x)$ is a point in $(\mathcal{R}^d, \mathcal{A})$. So a marked point process is a point process on the product space $(\mathcal{R}^d, \mathcal{A})$.

Consider a point process $\{N(B) : |B| \geq 0\}$ on a planar region $B \in \mathcal{R}^2$. $N(B)$ is the number of events on B. The events of interest in this model are the locations of the defects on the wafer. And the size of the defects is dealt with by the marked point process. A NHPP $N(B)$ is defined on a finite planar region B with the following postulates:

1. $N(B)$ has a Poisson distribution with mean $\mu(B) = \int_B \lambda(\mathbf{s}) d\mathbf{s}$, where $\mathbf{s} \in B$ and $\lambda(\mathbf{s})$ is the intensity function on B defined as equation (4.8).

2. Given $N(B) = n$, the n events form an independent random sample from the distribution on B with a PDF proportional to $\lambda(\mathbf{s})$.

The intensity function $\lambda(\mathbf{s})$ is defined as

$$\lambda(\mathbf{s}) \equiv \lim_{|d\mathbf{s}|\to 0} \frac{E[N(d\mathbf{s})]}{|d\mathbf{s}|}. \tag{4.8}$$

Thus, we can compute the average number of defects on B with

$$E[N(B)] = \int_B \lambda(\mathbf{s}) d\mathbf{s}.$$

IV.4. Yield models

IV.4.1. Point defect yield model

Let's consider a yield model based on the results shown above. In compound Poisson yield models, the defect variation is modeled by the defect density distribution which doesn't include spatial information about the defect density variation. The new yield models basically use the Poisson yield model, equation (4.4). However, the average number of faults varies die-by-die and this is computed using the intensity function of the spatial NHPP. Since the equation (4.3) is not identical for all dies, we need to modify the definition of yield as the ratio of the number of conforming products at the end of the production process to the number of potentially usable products[**?**]. The point defect model treats each defect as a fault, e.g., any defect located on a die area is assumed to cause a failure of the chip regardless of the location on the die or the size of the defect.

The following is the mathematical derivation of a new yield model. Define $\{B_j\}_1^k$ as a set of finite disjoint regions of the die area on the wafer, where k is the

number of dies on a wafer. Let's define a random variable

$$X_{B_j} = \begin{cases} 1 & \text{if } B_j \text{ is good,} \\ 0 & \text{if } B_j \text{ is defective, } j = 1, \cdots, k. \end{cases}$$

Then, the chip yield of die area B_j given $\lambda(\mathbf{s})$ is

$$\begin{aligned} E[X_{B_j}] &= \Pr\{X_{B_j} = 1\} \\ &= \Pr\{\text{no defects on die } B_j\} \\ &= P_0 \\ &= e^{-\mu(B_j)}, \quad j = 1, \cdots, k, \end{aligned} \tag{4.9}$$

where $\mu(B_j) = \int_{B_j} \lambda(\mathbf{s}) d\mathbf{s}$. The last equation holds by the first postulate of the spatial NHPP. If the intensity function is regarded as random with random coefficients $\boldsymbol{\beta}$ (for detailed discussion, refer to section IV.5.1), then the chip yield is

$$\begin{aligned} E[X_{B_j}] &= E[E[X_{B_j}|\boldsymbol{\beta}]] \\ &= \int_{\boldsymbol{\beta}} \exp\left\{ - \int_{B_j} \lambda(\mathbf{s}|\boldsymbol{\beta}) d\mathbf{s} \right\} f(\boldsymbol{\beta}) d\boldsymbol{\beta}, \end{aligned}$$

for $j = 1, \cdots, k$.

To build a yield model, let's define a new random variable $X = \sum_{i=1}^{k} X_{B_i}$. X represents the number of conforming dies on a wafer. Thus, we have a yield model defined as the average portion of conforming dies per wafer:

$$\begin{aligned} Y &= \frac{E[X]}{k} \\ &= \frac{E[X_{B_1}] + E[X_{B_2}] + \cdots + E[X_{B_k}]}{k} \end{aligned} \tag{4.10}$$

IV.4.2. General defect yield model

Given the location of a defect, a bigger defect is more likely to cause the failure of a chip. In order to deal with various defect sizes, marked point process is adopted. The size of defect x is the spatial marked random quantity of the marked point process. The intensity of the spatial marked point process on the product space $(\mathcal{R}^2 \times \mathcal{A})$ is defined:

$$\nu(\mathbf{s}, x) \equiv \lim_{|d\mathbf{s} \times dx| \to 0} \frac{E[N(d\mathbf{s} \times dx)]}{|d\mathbf{s} \times dx|} \tag{4.11}$$

Then $\int_B \nu(\mathbf{s}, dx) d\mathbf{s}$ is the average number of defects of size dx on region B and, consequently, the defect density of defect size x on region B is

$$D_0(x) = \frac{\int_B \nu(\mathbf{s}, x) d\mathbf{s}}{|B|}.$$

From equation (4.2), the average number of faults on region B, $\mu(B)$ is:

$$\begin{aligned} \mu(B) &= \int_0^\infty A_c(x) D_0(x) dx \\ &= \int_0^\infty A_c(x) \frac{\int_B \nu(\mathbf{s}, x) d\mathbf{s}}{|B|} dx. \end{aligned}$$

The chip yield of region B is

$$E[X_B] = \exp\left\{ - \int_0^\infty A_c(x) \frac{\int_B \nu(\mathbf{s}, x) d\mathbf{s}}{|B|} dx \right\}.$$

Note that the intensity function of the marked Poisson process is factorized as

$$\nu(\mathbf{s}, x) = \lambda(\mathbf{s}) g(x|\mathbf{s}),$$

where $g(x|\mathbf{s})$ is the conditional distribution of the defect size given the defect location.

The marginal defect size distribution on region B is

$$g(x) = \int_B g(x|\mathbf{s}) \frac{\lambda(\mathbf{s})}{\int_B \lambda(\mathbf{s}) d\mathbf{s}} d\mathbf{s}.$$

Then the critical area of chip region B is

$$\begin{aligned} A_c(B) &= \int_0^\infty A_c(x) g(x) dx \\ &= \int_0^\infty A_c(x) \int_B g(x|\mathbf{s}) \frac{\lambda(\mathbf{s})}{\int_B \lambda(\mathbf{s}) d\mathbf{s}} d\mathbf{s} dx \\ &= \frac{1}{\int_B \lambda(\mathbf{s}) d\mathbf{s}} \int_0^\infty A_c(x) \int_B g(x|\mathbf{s}) \lambda(\mathbf{s}) d\mathbf{s} dx. \end{aligned}$$

If $g(x|\mathbf{s}) = g(x)$, then

$$\begin{aligned} E[X_B] &= \exp\left\{ -\frac{\int_0^\infty \int_B A_c(x) g(x|\mathbf{s}) \lambda(\mathbf{s}) d\mathbf{s} dx}{|B|} \right\} \\ &= \exp\left\{ -\frac{\int_0^\infty A_c(x) g(x) dx \int_B \lambda(\mathbf{s}) d\mathbf{s}}{|B|} \right\}. \end{aligned}$$

If $\boldsymbol{\beta}$ is regarded as a random vector, then

$$E[X_B] = E[E[X_B|\boldsymbol{\beta}]] = \int_{\boldsymbol{\beta}} \exp\left\{ -\frac{\int_0^\infty A_c(x) g(x) dx \int_B \lambda(\mathbf{s}|\boldsymbol{\beta}) d\mathbf{s}}{|B|} \right\} f(\boldsymbol{\beta}) d\boldsymbol{\beta}.$$

The yield model is the same as equation (4.10).

IV.5. Intensity function

IV.5.1. Estimation

The nonhomegeneity of the defect density on a wafer is modeled by the intensity function. In this study, a parametric intensity function is considered. Assume an

intensity function:

$$\lambda(\mathbf{s}|\boldsymbol{\beta}) = q[\boldsymbol{\beta}^T \mathbf{h}(\mathbf{s})], \tag{4.12}$$

where $\mathbf{h}(\mathbf{s})$ is a known function of location; $q(\cdot)$ is a function and $\boldsymbol{\beta}$ represents the vector of the coefficients. The estimator of the unknown parameters $\boldsymbol{\beta}$ is obtained by maximizing the likelihood function [33]:

$$L(\boldsymbol{\beta}|\mathbf{s}) = \prod_{i=1}^{m} \lambda(\mathbf{s}_i|\boldsymbol{\beta}) \exp\left\{ - \int_A \lambda(\mathbf{s}|\boldsymbol{\beta}) d\mathbf{s} \right\}, \tag{4.13}$$

where m is the number of defects on a wafer, $\mathbf{s}_i \in A$ is the location of defect i, $i = 1, \cdots, m$ and A is the wafer region. Generally, there is no closed form for the maximum likelihood estimator of $\boldsymbol{\beta}$.

Other than the nonhomogeneity of intensity related to the location on the wafer, variations in the intensity function among wafers should be taken into account in estimation. To this end, I consider a doubly stochastic point process(or Cox process). A Cox process is defined as a point process N with random intensity function $\Lambda(\mathbf{s})$. If $\Lambda = \lambda$, then N is a NHPP with $\lambda(\mathbf{s})$. A hierarchical model which has random coefficients is a general assumption for a random intensity function. One wafer is considered as a realization from the random defect generation mechanism. The random intensity function is defined as equation (4.12) with random vector $\boldsymbol{\beta}$ that follows a multivariate distribution, $f(\boldsymbol{\beta}|\boldsymbol{\theta})$. The estimation method for $\boldsymbol{\theta}$ is as follows:

1. Collect M wafers.
2. Get $\hat{\boldsymbol{\beta}}_1, \cdots, \hat{\boldsymbol{\beta}}_M$ using maximum likelihood estimation with equation (4.13).
3. Regard $\hat{\boldsymbol{\beta}}_1, \cdots, \hat{\boldsymbol{\beta}}_M$ as a random sample of size M from a multivariate distribution. And estimate $\boldsymbol{\theta}$.

For the intensity function of a nonhomogeneous marked Poisson process, let $\boldsymbol{\alpha}$ be the vector of parameters of $g(x|\mathbf{s})$; then the loglikelihood function is

$$\begin{aligned} l(\boldsymbol{\alpha},\boldsymbol{\beta}|\mathbf{s},\mathbf{x}) &= \sum_{i=1}^{m} \log \nu(\mathbf{s}_i, x_i|\boldsymbol{\alpha},\boldsymbol{\beta}) - \int_0^{\infty} \int_A \nu(\mathbf{s}, x|\boldsymbol{\alpha},\boldsymbol{\beta}) d\mathbf{s} dx \\ &= \sum_{i=1}^{m} \log \lambda(\mathbf{s}_i|\boldsymbol{\beta}) - \int_A \lambda(\mathbf{s}|\boldsymbol{\beta}) d\mathbf{s} + \sum_{i=1}^{m} \log g(x_i|\mathbf{s}_i, \boldsymbol{\alpha}). \end{aligned}$$

If the first two terms and the third term have no common parameters, we can maximize the loglikelihood by maximizing them separately.

A fixed intensity function that represents the defect pattern of a manufacturing process can be obtained from $\Lambda(\mathbf{s})$ by taking $E[\Lambda(\mathbf{s})]$. Also, we may use

$$\lambda(\mathbf{s}) = q\{E[\boldsymbol{\beta}]^T \mathbf{h}(\mathbf{s})\} \tag{4.14}$$

in order to reduce the effort in computation if the variance of $f(\boldsymbol{\beta})$ is small.

IV.5.2. Construction

A defect is a variation in quality that may cause the failure of the circuit. A visible defect is a visible object on the wafer such as a particle, hole, etc. The causes of these visible defects are particles in the cleanroom, particles from equipments, chemical stains or scratches on the wafer due to improper batch handling. The pattern of defects on the wafer results from the superposition of results from the two defect generation mechanisms. One is the random cause that creates the global defect pattern. The other is the assignable cause that generates the local patterns that are normally found as members of a defect cluster. The global pattern results from cleanroom environment variations such as particles in cleanroom and thermal variation in the annealing process. The local pattern results from local destructive

mechanism such as the mishandling of wafers, particles from equipment, dislocation of wafer crystals or chemical stains on the wafers.

In order to model different sources of the defect generation mechanisms in an intensity function, I take the advantage of the additivity of the intensity function. Since the intensity function of a NHPP contains the defect density variation by location, we can add up all of the intensity functions estimated for the different sources of defects. If we define $\lambda_k(\mathbf{s})$ to be the intensity function of the k^{th} defect source, then the overall intensity function is

$$\lambda(\mathbf{s}) = \sum_{k=1}^{G} \lambda_k(\mathbf{s}),$$

where G is the number of different sources of defects. Estimating the individual intensity function of each defect generation mechanism can be accomplished by a method called clustering.

Clustering is a procedure that categorizes data into meaningful subgroups. In model-based clustering, the observations are assumed to be generated by a mixture probability distribution:

$$f(\mathbf{s}|\boldsymbol{\beta}) = \sum_{k=1}^{G} p_k f_k(\mathbf{s}|\boldsymbol{\beta}_k),$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_G)$, $p_k \geq 0 \; \forall k$, $\sum_{k=1}^{G} p_k = 1$. $f_k(\mathbf{s}|\boldsymbol{\beta}_k)$ is the probability density at location $\mathbf{s}$ from the k^{th} component and is obtained from $\lambda_k(\mathbf{s})$ by

$$f_k(\mathbf{s}|\boldsymbol{\beta}_k) = \frac{\lambda_k(\mathbf{s}|\boldsymbol{\beta}_k)}{\int_A \lambda_k(\mathbf{s}|\boldsymbol{\beta}_k) d\mathbf{s}}.$$

The characteristics of clusters can be summarized by the parameters such as $p_k, \boldsymbol{\beta}_k$, $k = 1, \cdots, G$.

In the estimation of the parameters, the locations are regarded as missing data. The complete data are considered to be $\mathbf{y}_i = (\mathbf{s}_i, \mathbf{z}_i)$, where $\mathbf{z}_i = (z_{i1}, \cdots, z_{iG})$,

where

$$z_{ik} = \begin{cases} 1, \text{ if } \mathbf{s}_i \text{ belongs to } k^{th} \text{ cluster,} \\ 0, \text{ otherwise.} \end{cases}$$

$\mathbf{z}_i$ is assumed to follow the identical and independent multinomial distribution of one trial with parameters $p_1, \cdots, p_G$. And the distribution of $\mathbf{s}_i$, given $\mathbf{z}_i$, is given by

$$f(\mathbf{s}_i) = \prod_{k=1}^{G} f_k(\mathbf{s}_i)^{z_{ik}}.$$

Using the complete data, the estimators are obtained by maximizing the complete-data likelihood function [34, 35, 36]:

$$L(\boldsymbol{\beta}, \mathbf{p}, \mathbf{z}|\mathbf{s}) = \prod_{i=1}^{n} \prod_{k=1}^{G} [p_k f_k(\mathbf{s}_i|\boldsymbol{\beta}_k)]^{z_{ik}}. \tag{4.15}$$

Even with the clustering algorithm, the maximization problem is a combinatorial optimization problem and is hard to solve. An iterative solving method is provided by the EM(expectation-maximization) algorithm that uses iterative steps of expectation and maximization. The EM algorithm is a general algorithm for computing the maximum likelihood estimates of $p_k, \boldsymbol{\beta}_k$ $(1 \leq k \leq G)$. The EM algorithm with complete-data likelihood criteria is as follows:

Repeat step 1 and 2

1. M-step

Maximize the complete likelihood function, equation(5.1), in terms of $\mathbf{p}$ *and* $\boldsymbol{\beta}$ *given* $\hat{\mathbf{z}}^{(m-1)}$*, the conditional expectation of* $(m-1)^{st}$ *iteration.*

The optimal solution at the m^{th} *iteration is*

$$\hat{p}_k^{(m)} = \frac{\sum_{i=1}^{n} \hat{z}_{ik}^{(m-1)}}{n}$$

and $\hat{\boldsymbol{\beta}}_k^{(m)}$ *is the optimal solution of the likelihood function*

$$L(\boldsymbol{\beta}_k|\mathbf{s}) = \prod_{i=1}^{n} [\lambda_k(\mathbf{s}_i|\boldsymbol{\beta}_k)]^{\hat{z}_{ik}^{(m-1)}} \exp\left\{-\int_A \lambda_k(\mathbf{s}|\boldsymbol{\beta}_k)d\mathbf{s}\right\},$$

where A is the area of interest. In this case it is the region of the wafer.

2. E-step

Compute the conditional expectation $E[z_{ik}|\mathbf{s}_i, \boldsymbol{\beta}]$*, the conditional probability that* $\mathbf{s}_i$ *is from the* k^{th} *cluster.*

$$\hat{z}_{ik}^{(m)} = \frac{\hat{p}_k^{(m)} f_k(\mathbf{s}_i|\hat{\boldsymbol{\beta}}_k^{(m)})}{\sum_{j=1}^{G} \hat{p}_j^{(m)} f_j(\mathbf{s}_i|\hat{\boldsymbol{\beta}}_j^{(m)})}.$$

until the convergence criteria are satisfied.

If the intensity function of a local defect has the form of $\lambda(\mathbf{s}) = w\phi(\mathbf{s})$, where $\phi(\cdot)$ is a probability density function, then the loglikelihood function given $\mathbf{z}$ is

$$\begin{aligned} l(w, \boldsymbol{\beta}|\mathbf{s}) &= \log\left\{\prod_{i=1}^{n} [w\phi(\mathbf{s}_i|\boldsymbol{\beta})]^{z_i} e^{-\int_A w\phi(\mathbf{s}|\boldsymbol{\beta})d\mathbf{s}}\right\} \\ &= (\log w)\sum_{i=1}^{n} z_i + \sum_{i=1}^{n} z_i \log\phi(\mathbf{s}_i|\boldsymbol{\beta}) - w\int_A \phi(\mathbf{s}|\boldsymbol{\beta})d\mathbf{s}. \end{aligned} \tag{4.16}$$

Let $(w^*, \boldsymbol{\beta}^*)$ be the optimal solution of the maximization problem. If we assume that the optimal solution is not obtained on the boundary of the feasible region, then by the necessary condition for an optimum,

$$\left.\frac{\partial l(w, \boldsymbol{\beta}|\mathbf{s})}{\partial w}\right|_{w^*, \boldsymbol{\beta}^*} = \frac{\sum_{i=1}^{n} z_i}{w^*} - \int_A \phi(\mathbf{s}|\boldsymbol{\beta}^*)d\mathbf{s} = 0.$$

Thus,

$$w^* \int_A \phi(\mathbf{s}|\boldsymbol{\beta}^*)d\mathbf{s} = \sum_{i=1}^{n} z_i.$$

Now the equation (4.16) can be considered as a separable problem and the optimal solution $\boldsymbol{\beta}^*$ is obtained from MLE of $\phi(\mathbf{s})$ given $\mathbf{z}, \mathbf{s}$ and $w^* = \sum_{i=1}^{n} z_i / \int_A \phi(\mathbf{s}|\boldsymbol{\beta}^*) d\mathbf{s}$.

IV.6. Simulation

In this section, simulation results for the point defect yield model and the general defect yield model are provided. For the point defect yield model, the effects of various coefficients, the die size and the existence of local defects are investigated. For the general defect yield model, the effect of the various parameters of the defect size distribution and the critical area are examined. Two die sizes of 1×1 cm^2 and 1.5×1.5 cm^2 are used on a 20cm wafer, which make 260 and 112 dies, respectively.

IV.6.1. Simulation setup

Relative error(RA), $\frac{|Y-\hat{Y}|}{Y}$ is used as a performance measure of the yield models, where Y is the counted yield (which is counted from the wafers) and $\hat{Y}$ is the estimated yield (which is estimated with the new yield models). The estimated yield is computed with an intensity function obtained using equation (4.14). To evaluate the average performance and the stability of the models, average relative error(ARE) and maximum relative error(MRE) are used. They are simply the average and maximum values of the RA's from runs of the simulations. 30 wafers are used for computing counted yield and estimated yield.

As discussed in section IV.5.2, I consider two types of defects, global defects and local defects. The global defects are generated by the thinning method [37]; the local defects are generated from the bivariate normal distribution. Figure 7 shows the simulation patterns of those defect types.

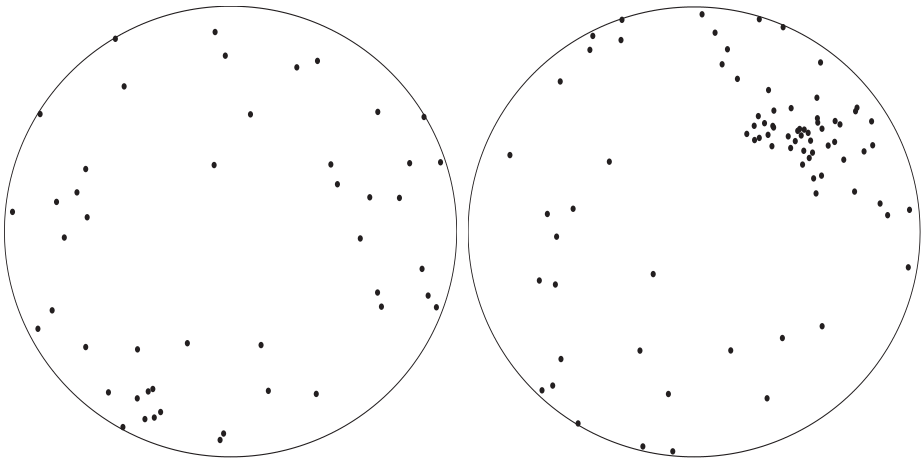


Figure 7-(a) Figure 7-(b)

Figure 7: Simulated defects.
(a) is the global defect pattern and (b) is the superposition of global and local patterns.

For the intensity function of the global defects, a quadratic function form is used. The advantage of the quadratic intensity function is that the constant term can be considered as the intensity of the background defect pattern that generates the random defects and the other terms as the intensity of the edge defects that produces the ring shape defect pattern on the wafers. Polar coordinate is used with the assumption that the defect pattern is isotropic, i.e., there is no angular term in the intensity function. The intensity function used is

$$\lambda_1(r|\boldsymbol{\beta}) = \frac{1}{10^3}(\beta_0 + \beta_1 r + \beta_2 r^2),$$

where r represents the distance from the center of a wafer in centimeters. We scale the intensity with 10^3 to deal with the coefficients, β's, with convenient numbers between 0 and 10. The coefficients are assumed to be independent and to follow a

left truncated normal distribution with $\mu,\ \sigma$ and truncation point k:

$$f(y) = \frac{\frac{1}{\sigma}\phi(\frac{y-\mu}{\sigma})}{1-\Phi(\frac{k-\mu}{\sigma})},$$

where ϕ and the Φ are PDF and CDF of the standard normal distribution, respectively. The truncation point k is set to be zero to guarantee nonnegativity of the coefficients. To examine the effect of different clustering effect with various variances of $\boldsymbol{\beta}$, we use two mean vectors ($\mu_{\beta_1}\ \mu_{\beta_2}\ \mu_{\beta_3}$) and three variance vectors ($\sigma^2_{\beta_1}\ \sigma^2_{\beta_2}\ \sigma^2_{\beta_3}$).

For the intensity function of local defects, we use the bivariate normal distribution:

$$\lambda_2(\mathbf{s}|w, \boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N) = w\frac{1}{2\pi|\boldsymbol{\Sigma}_N|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\mathbf{s}-\boldsymbol{\mu}_N)^T\boldsymbol{\Sigma}_N^{-1}(\mathbf{s}-\boldsymbol{\mu}_N)\right\},$$

where $w \geq 0$ is a parameter that governs the height of the intensity function. The multivariate normal distribution is widely considered in model-based clustering [38, 39], since the properties of a defect cluster can be described by the parameters of the distribution- $\boldsymbol{\mu}_N$ represents the center of the cluster and $\boldsymbol{\Sigma}_N$ describes the shape and orientation of the defect cluster. Like the intensity function of the global defects, the parameters of the bivariate normal distribution, $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N$, are assumed to be random.

The defect size distribution is assumed as

$$g(x) = \begin{cases} \frac{2(p-1)x}{(p+1)x_0^2}, & \text{if } x \leq x_0, \\ \frac{2(p-1)x_0^{p-1}}{(p+1)x^p}, & \text{if } x > x_0, \end{cases}$$

with $p = 3$ [11], and graph of the distribution function is illustrated in Figure 1. The number of defects and the defect locations are generated with the same algorithms

as used for the point defect yield model. And the critical area model is assumed as

$$A_c(x) = \begin{cases} 0, & \text{if } x < w_0, \\ \frac{x-w_0}{w_1-w_0}|B|, & \text{if } w_0 \le x < w_1, \\ |B|, & \text{if } x \ge w_1. \end{cases}$$

Figure 8 sketches the critical area with this model. We assume that the critical

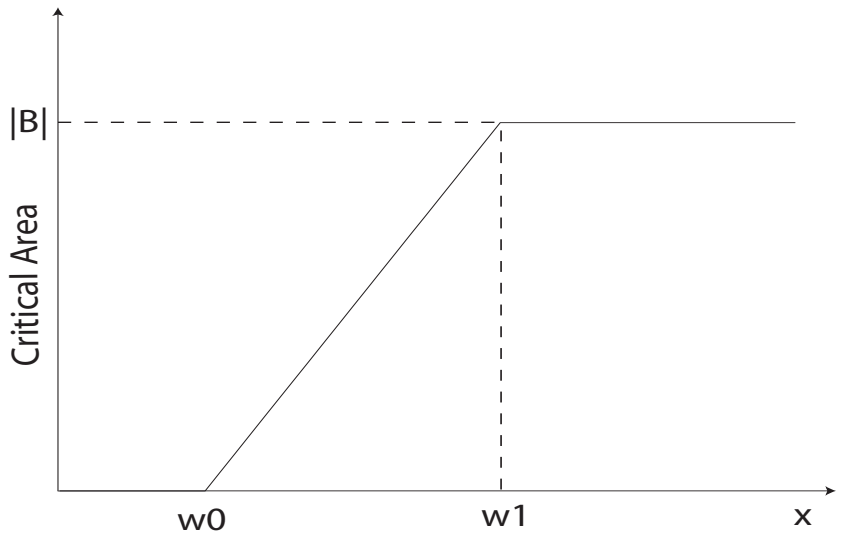


Figure 8: Critical area.

area is linear to the defect size x, such that $w_0 \le x \le w_1$, and is bounded by the die area. It is an useful assumption that the maximum critical area is equal to the die area since the probability density of a large defect size is very small [11].

IV.6.2. Simulation results

The simulation of the point defect yield model with global defects is summarized in Table 2. As shown in the table, ARE's that are less than 0.5% indicate the accuracy of the yield model and MRE's that are less than 2% show that the point defect model

and estimation method are stable. The ARE and MRE remain stable as the variance of the parameters increases but they change as the die size increases.

Table 2: Simulation with various intensity functions

Die size	$(\mu_{\beta_1}\ \mu_{\beta_2}\ \mu_{\beta_3})$	$(\sigma^2_{\beta_1}\ \sigma^2_{\beta_2}\ \sigma^2_{\beta_3})$	ARE	MRE
1×1	(6 6 2)	(1 1 0.1)	0.001702	0.005125
		(5 5 0.5)	0.001912	0.005175
		(10 10 1)	0.002274	0.006597
1×1	(2 2 3)	(1 1 0.1)	0.002350	0.007143
		(5 5 0.5)	0.002658	0.007280
		(10 10 1)	0.002609	0.008993
1.5×1.5	(6 6 2)	(1 1 0.1)	0.004379	0.012733
		(5 5 0.5)	0.003925	0.015019
		(10 10 1)	0.004139	0.012277
1.5×1.5	(2 2 3)	(1 1 0.1)	0.004992	0.014548
		(5 5 0.5)	0.004665	0.014916
		(10 10 1)	0.004878	0.019051

Table 3: Simulation with or without clustering algorithm

Local defects	Without clustering		With clustering	
	ARE	MRE	ARE	MRE
All over wafer	0.019236	0.035538	0.028683	0.287910
On local area	0.014556	0.029932	0.005447	0.024954

Table 3 presents the results of simulations of the point defect yield model with global and local defects. The global defects are generated from the intensity function with $(\mu_{\beta_1}\ \mu_{\beta_2}\ \mu_{\beta_3}) = (2\ 2\ 3)$ and $(\sigma^2_{\beta_1}\ \sigma^2_{\beta_2}\ \sigma^2_{\beta_3}) = (1\ 1\ 0.1)$. Forty local defects are generated from the bivariate normal distribution for each wafer. For

the simulation with the "all over wafer," the mean vector of the bivariate normal distribution, $\boldsymbol{\mu}_N$, is assumed to be distributed uniformly on the wafer. The results show that if the local defect clusters occur all over the wafer area, the intensity function just with global effects is still able to estimate yield with high accuracy. Introducing a clustering algorithm for estimation makes it even worse. The random occurrence of local defects shifts the intensity function of the global defects up, and the intensity function of the local defects provides no information on the defects. In the simulation with the "on local area," the $\boldsymbol{\mu}_N$ is assumed to follow the normal distribution. Table 3 indicates that the performance of the yield model is improved by introducing the clustering algorithm.

Table 4: Simulation of the general defect yield model with various defect size distributions and critical area functions

w_0(micron)	w_1(micron)	x_0(micron)	ARE	MRE
0.1	0.4	0.3	0.004102	0.019383
0.1	0.4	0.5	0.004190	0.014005
0.1	1	0.3	0.002697	0.010026
0.2	0.4	0.3	0.004353	0.022170

For the simulation of the general defect yield model, only global defects are used with a 1.5×1.5 cm^2 die size. The global defects are generated from the intensity function with $(\mu_{\beta_1}\ \mu_{\beta_2}\ \mu_{\beta_3}) = (2\ 2\ 3)$ and $(\sigma^2_{\beta_1}\ \sigma^2_{\beta_2}\ \sigma^2_{\beta_3}) = (1\ 1\ 0.1)$. The results of simulations with various w_0, w_1 and x_0 are summarized in Table 4. All of the results show the accuracy and stability of the model.

IV.7. Conclusion

I build new yield models of ICs that incorporate the spatial defect distributions of wafers based on spatial NHPP, which has never before been successfully modeled. The intensity function of the spatial NHPP, combined with the model-based clustering method is flexible enough to deal with various defect sizes and defect patterns including local defects. Judging from the simulation studies, the accuracy of the yield estimation is very high. However, the real benefits of the new models can be found in the fact that the information about the defect distributions also provides appropriate measures for fabrication process control since the information about spatial defect distributions can be used to (1) visualize the defect pattern through the intensify function, (2) incorporate various intensity functions estimated for different defect types or for different manufacturing processes, (3) reduce the wafer map information that is critical for process control with several parameters of the intensity functions such as location and shape of local defect clusters and global defect patterns, and (4) find sources of yield loss by identifying local defects generated from assignable causes.

CHAPTER V

MODEL-BASED CLUSTERING FOR INTEGRATED CIRCUIT YIELD ENHANCEMENT

This chapter studies the defect data analysis method for semiconductor yield enhancement. Given the defect locations on a wafer, the local defects generated from the assignable causes are classified from the global defects generated from the random causes by model-based clustering, and the clustering methods can identify the characteristics of local defect clusters. The information obtained from this method can facilitate process control, particularly, root-cause analysis. The global defects are modeled by the spatial nonhomogeneous Poisson process, and the local defects are modeled by the bivariate normal distribution or by the principal curve.

V.1. Introduction

Detecting process variation, the most serious factor affecting yield, is the starting point for yield improvement in IC products. In semiconductor fabrication, which has more than two hundred process steps, and each of which has more than ten key process measures, tracing process variation by monitoring those measures is not well developed nor efficient [2]. Instead, wafer map analysis is an useful method for controlling process variation. Wafer map analysis is a procedure of extracting information about process variation from the visible defects. A defect is a variation in quality that may cause circuit failure. And visible defects are the visible objects on a wafer, such as particles, extra material, holes, etc, and they are the main sources of information about process variation.

Wafer map analysis can be classified largely into visual inspection or auto-

mated defect scanning. Visual inspection with a high-resolution microscope is the main method employed by defect data root-cause analysis. This method has the advantages of low cost and root-cause analysis capability, and it is the only method that can detect the causes of defects. By examining size, location, color and shape with a high resolution microscope, the visual inspector finds the defect cause. However, it also has limitations- the speed of visual inspection is so low that it takes several hours to examine a wafer; it takes a long time to train visual inspectors; and visual inspectors cannot concentrate for very long time periods due to fatigue. Thus, considering the huge number of wafers produced, examining all of the defects is far too time consuming and costly [4]. On the other hand, low-resolution automated defect scanning uses laser light to scan the surface of wafers. Using scattered light, it can identify the locations of defects and their relative sizes. It is fast enough to examine a wafer in several minutes. However, it cannot replace visual inspection for root-cause analysis because of its low resolution. Thus, root-cause analysis relies mostly on visual inspection which is slow and inconsistent.

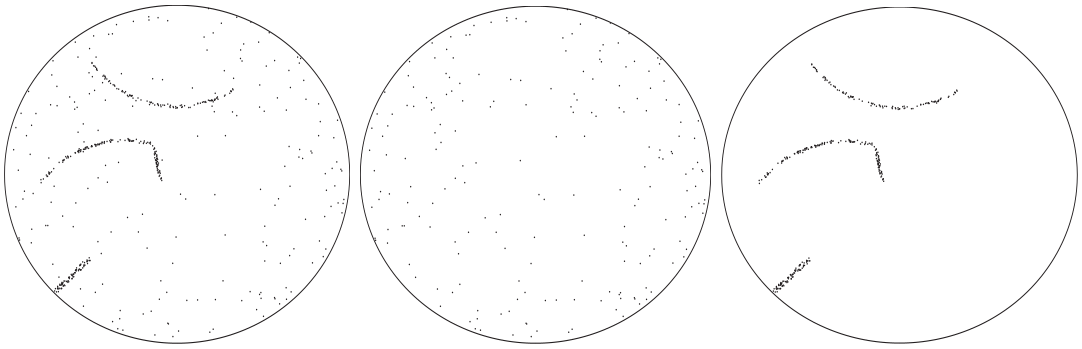
There is another factor that affects the profit of a semiconductor company as much as improved yield. That is the time factor of the yield learning process. The yield learning rate, i.e., how soon a process reaches the desired yield level, is as crucial as the yield itself to the success of the factory [3]. Therefore, a technique filling the gap between the above two wafer map inspection methods is needed in order to sensitize process control to process variations and to shorten the root-cause analysis cycle. Also, although eliminating all of the variation sources is the ideal way to improve yield, it is not the optimal way from an economical perspective. The sources of variation can be categorized into random causes and assignable causes. Since the cost of fixing random causes is much higher than the cost of fixing assignable causes, yield improving efforts are mostly focused on finding and removing assignable causes.

Consequently, we need a method that can discover the assignable causes from the wafer map.

The purpose of this study is to provide an automated wafer map analysis method that combines the strengths of the two wafer inspection methods to detect the existence, location and orientation of defect clusters in order to distinguish the wafers that deserve further visual investigation. Root-cause analysis is still something of an art in that it requires subjective judgements. However, it can be facilitated to produce better results by the reduction and preprocessing of data. The basic idea of this study is to use spatial features like cluster patterns and the geometrical properties of these clusters to identify the defects generated by the assignable causes from low resolution defect data and to feed the results to high resolution analysis.

The defect pattern on a wafer results from the superposition of two defect patterns as shown in Figure 9. One is the global defect pattern which results from random causes. These random causes produce defects on ICs by chance. The random causes are cleanroom environment variations such as particles in the cleanroom and thermal variation in the annealing process, and they are expensive to remove. The other pattern is local defects which are generated by assignable causes. The assignable causes generate defects systematically, and thereby, generate a set of aggregated defects, a cluster. The assignable causes are local destructive mechanisms, such as the mishandling of wafers, particles from equipment, dislocation of wafer crystals or chemical stains on the wafers. Thus, the defects from assignable causes can be identified by detecting the presence of defect clusters and partitioning the defect clusters from the global defects. Each defect cluster can be categorized into its defect generation causes according to its spatial pattern. For example, a cluster of curvilinear shape is assigned to scratches, and a round shape cluster is assigned to particles from equipment or chemical stains. Using the spatial pattern of the local

defect cluster, valuable information about the defect cause can be obtained for the visual inspection.



(a) (b) (c)

Figure 9: Simulated defect patterns.
Figure (a) is a stacked defect pattern of different sources of defect generation mechanisms: Global defects (b) and three local defect clusters (c).
Figure (c) shows curvilinear, double curvilinear and ellipsoidal pattern.

There have been numerous studies on retrieving information about process variations from defect data or test data [40, 41, 42, 43, 44, 20, 45]. Traditionally, summary measures such as wafer yield or lot yield were used to this end. But due to the presence of defect clustering, these measures are not appropriate for extracting useful information about process variations. Methods using probe data of the locations of defective chips are discussed in the literature [20, 44]. But these methods are more for detecting the existence of process variations than for root-cause analysis. Spatial pattern recognition using wafer probe data on the location of defective chips is studied in literature [40, 45]. The results provide information about the fabrication process and can be used for root-cause analysis. However, the judgements of root-cause analysis can be made only from defect data [42]. This is because multiple defects on a chip result in a single failed chip, and, consequently, lots of information

is lost. Gleason et al. [43] employ an automated clustering algorithm using artificial intelligence. Chen and Liu [41] use neural-networks for pattern recognition, and Cunningham and Mackinnon [42] use an empirical clustering algorithm. These methods are focused on retrieving the spatial features of the defects. This study extends these works to a combined yield management method that simultaneously identifies defect clusters and obtains information for the yield model.

I propose a model-based clustering for wafer map analysis. The model-based clustering has the following advantages: (1) It is so flexible no training data is needed, and new defect patterns can be easily detected, and (2) the results of the clustering provide information for yield estimation as well as process monitoring. Probabilistic methods can apply the Bayesian information criteria(BIC) to estimate the number of clusters, and they are capable of partitioning the overlapped clusters. In model-based clustering, the data are considered to be generated from a mixture distribution, and normally, clustering with the multivariate normal distribution or detecting clusters from random background noise are studied. However, curvilinear defect patterns are observed on the defect wafer map, and the background noise is not random. Thus, I propose a model-based clustering method that uses a distribution estimated from the global defects with a spatial nonhomogeneous Poisson process and the principal curve to model the curvilinear defect patterns. According to the pattern of the local defects, they can be modeled with bivariate normal distributions or the principal curves.

The rest of this chapter consists of the following. Section V.2 reviews the background of model-based clustering. In this section, we review probability models for model-based clustering, the maximization criteria for the parameter estimation, expectation-maximization(EM) and classification EM(CEM) algorithm for the maximization problem, BIC for estimating the number of clusters and the principal curve.

In Section V.3, the clustering strategy for wafer defect data analysis is discussed. Section V.4 describes the simulated defect generation and discusses the results of clustering. And Section V.5 concludes this chapter.

V.2. Model-based clustering

V.2.1. Probability models

My interest in this study is in classifying the defects into clusters and gaining information about the clusters. A cluster is defined as an aggregation of the defects. However, it can be redefined as the meaningful subgroup of defects that are generated from the same defect generation mechanism. Thus, the global defects that occur all over the wafer are also called a cluster. It is assumed that the global defects are generated from random causes. Although there might be several sources of variation in these random causes, they are considered to be the same defect generation mechanism. The purpose of this study is to discriminate the defects from assignable causes and to find the characteristics of their clusters using model-based clustering. The underlying assumptions are that the defects from assignable causes tend to cluster and that different defect generation mechanisms generate different defect patterns accordingly.

In model-based clustering, the observations are assumed to be generated by a mixture distribution:

$$f(\mathbf{s}|\boldsymbol{\theta}) = \sum_{k=1}^{G} p_k f_k(\mathbf{s}|\boldsymbol{\theta}_k),$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_G)$, $p_k \geq 0 \ \forall k$, $\sum_{k=1}^{G} p_k = 1$ and G is the number of components in the mixture. Without loss of generality, let $k = 1$ represent the cluster of global defects. The clusters generated from assignable causes are classified into different

clusters according to their causes. $f_k(\mathbf{s}_i|\boldsymbol{\theta}_k)$ is the probability density function(PDF) of an observation $\mathbf{s}_i$ from the k^{th} component. The characteristics of the clusters can be summarized by parameters such as p_k and $\boldsymbol{\theta}_k$, $k = 1, \cdots, G$. When it is uniformly distributed over the region of interest, the PDF of the cluster of global defects, $f_1(\mathbf{s}|\boldsymbol{\theta}_1)$, is [34, 36]

$$f_1(\mathbf{s}|\boldsymbol{\theta}_1) = \frac{1}{D},$$

where D is the area of the region of interest. For the clusters of local defects, the multivariate normal distribution has been widely considered in model-based clustering [38, 39]. The spatial defect pattern can be nicely described by the parameters of the distribution: the mean vector, $\boldsymbol{\mu}$, represents the center of the cluster, and the covariance matrix, $\boldsymbol{\Sigma}$, describes the shape and the orientation of the ellipsoid composed of the members of the cluster. A symmetric matrix, $\boldsymbol{\Sigma}$, has positive eigenvalues, and the largest eigenvalue and the corresponding eigenvector represent the magnitude and the direction of the main axis of the set of defects in multidimensional space, respectively. In this study, the defects are assumed to take two dimensions. Thus, the eigenvector which corresponds to the larger eigenvector represents the longer axis of the ellipsoid, and the ratio of the eigenvalues are equal to the ratio of the lengths of the two axis of the ellipsoid. The principal components of two dimensional data set are illustrated in Figure 10.

In the parameter estimation, the locations are regarded as missing data. The complete data are considered to be $\mathbf{y}_i = (\mathbf{s}_i, \mathbf{z}_i)$, where $\mathbf{z}_i = (z_{i1}, \cdots, z_{iG})$, where

$$z_{ik} = \begin{cases} 1, \text{ if } \mathbf{s}_i \text{ belongs to } k^{th} \text{ cluster,} \\ 0, \text{ otherwise.} \end{cases}$$

It is assumed that $\mathbf{z}_i$ follows an identical and independent multinomial distribution

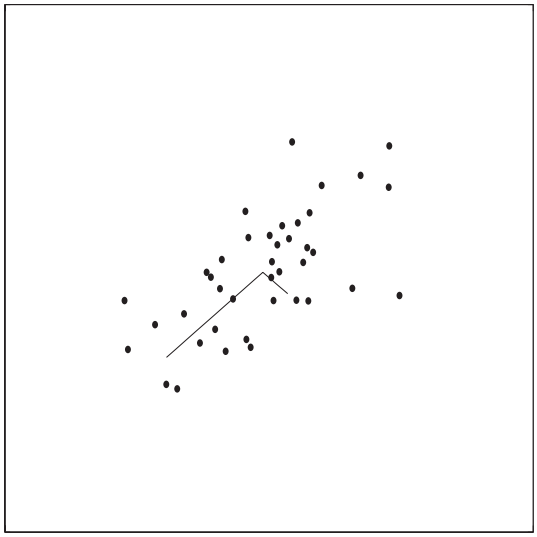


Figure 10: Principal components.

of a single trial with parameters $(p_1, \cdots, p_G)$ and that the PDF of $\mathbf{s}_i$ given $\mathbf{z}_i$ is given by

$$f(\mathbf{s}_i) = \prod_{k=1}^{G} f_k(\mathbf{s}_i)^{z_{ik}}.$$

Using the complete data, the estimators are obtained by maximizing one of the following likelihood criteria. The complete-data likelihood function is [34, 35, 36]

$$L_1(\boldsymbol{\theta}, \mathbf{p}, \mathbf{z}|\mathbf{s}) = \prod_{i=1}^{n} \prod_{k=1}^{G} [p_k f_k(\mathbf{s}_i|\boldsymbol{\theta}_k)]^{z_{ik}}. \tag{5.1}$$

The classification likelihood function is [39]

$$L_2(\boldsymbol{\theta}, \gamma_1, \cdots, \gamma_n|\mathbf{s}) = \prod_{i=1}^{n} f_{\gamma_i}(\mathbf{s}_i|\boldsymbol{\theta}_{\gamma_i}),$$

where the γ_i's are the indexing values for classification such that $\gamma_i = k$ if $z_{ik} = 1$.

V.2.2. EM algorithm for clustering

Maximization problems with L_1 or L_2 are hard to solve because of their combinatorial feature. An iterative solving method is provided by the EM(expectation-maximization) algorithm or by the CEM(classification EM) algorithm that use iterative steps for classification (in CEM), expectation and maximization. The EM algorithm with the L_1, equation (5.1) is studied in this section.

The EM algorithm is a general algorithm for computing the maximum likelihood estimates of p_k and $\boldsymbol{\theta}_k$ $(1 \leq k \leq G)$. The EM algorithm with the complete-data likelihood criteria is summarized as follows [46].

Repeat step 1 and 2

1. M-step

Maximize L_1 in equation (5.1) with respect to $\mathbf{p}$ and $\boldsymbol{\theta}$ given $\hat{\mathbf{z}}^{(m-1)}$, the estimate of the conditional expectation $E[z_{ik}|\mathbf{s}_i, \mathbf{p}, \boldsymbol{\theta}]$ of the $(m-1)^{st}$ iteration.

As a result of the maximization, the optimal solution at the m^{th} iteration is

$$\hat{p}_k^{(m)} = \frac{\sum_{i=1}^{n} \hat{z}_{ik}^{(m-1)}}{n}$$

and $\hat{\boldsymbol{\theta}}_k^{(m)}$ is the optimal solution of the likelihood function

$$L(\boldsymbol{\theta}_k^{(m)}|\mathbf{s}) = \prod_{i=1}^{n} [f_k(\mathbf{s}_i|\boldsymbol{\theta}_k^{(m)})]^{\hat{z}_{ik}^{(m-1)}}.$$

2. E-step

Compute $\hat{\mathbf{z}}^{(m)}$ given $\hat{\mathbf{p}}^{(m)}, \hat{\boldsymbol{\theta}}^{(m)}$:

$$\hat{z}_{ik}^{(m)} = \frac{\hat{p}_k^{(m)} f_k(\mathbf{s}_i|\hat{\boldsymbol{\theta}}_k^{(m)})}{\sum_{j=1}^{G} \hat{p}_j^{(m)} f_j(\mathbf{s}_i|\hat{\boldsymbol{\theta}}_j^{(m)})}.$$

until the convergence criteria are satisfied.

The initial point, $\hat{\mathbf{z}}^{(0)}$, is chosen arbitrarily or by certain rules. Since the algorithm depends on the initial point, the algorithm can be applied with several initial points, and the best result in terms of likelihood can be chosen [47].

V.2.3. CEM algorithm

The CEM algorithm is summarized [47]:

Repeat steps 1, 2 and 3

1. E-step

Compute $\hat{\mathbf{z}}^{(m)}$ given $\hat{\mathbf{p}}^{(m-1)}, \hat{\boldsymbol{\theta}}^{(m-1)}$:

$$\hat{z}_{ik}^{(m)} = \frac{\hat{p}_k^{(m-1)} f_k(\mathbf{s}_i|\hat{\boldsymbol{\theta}}_k^{(m-1)})}{\sum_{j=1}^{G} \hat{p}_j^{(m-1)} f_j(\mathbf{s}_i|\hat{\boldsymbol{\theta}}_j^{(m-1)})}.$$

2. C-step

Partition $\mathbf{s}_1, \cdots, \mathbf{s}_n$ into the sets of elements, $P_1^{(m)}, \ldots, P_G^{(m)}$, according to $\hat{z}_{ik}^{(m)}$. $\mathbf{s}_i$ will be partitioned in a cluster which provides the maximum $\hat{z}_{ik}^{(m)}$.

3. M-step

Compute the maximum likelihood estimates $\hat{p}_k^{(m)}, \hat{\boldsymbol{\theta}}_k^{(m)}$:

$$\hat{p}_k^{(m)} = \frac{\#P_k^{(m)}}{n},$$

where $\#P_k^{(m)}$ is the number of elements in $P_k^{(m)}$. $\hat{\boldsymbol{\theta}}_k^{(m)}$ is obtained by maximizing the likelihood

$$L(\boldsymbol{\theta}_k^{(m)}|\mathbf{s}) = \prod_{\mathbf{s}_i \in P_k^{(m)}} [f_k(\mathbf{s}_i|\boldsymbol{\theta}_k^{(m)})].$$

until the convergence criteria are satisfied.

V.2.4. Number of clusters

The number of the clusters in the defect data is estimated by the BIC [48]. The applications of BIC in model-based clustering can be found in [34, 35]. The BIC is

$$2 \log p(\mathbf{s}|\mathcal{M}) + C,$$

where $p(\mathbf{s}|\mathcal{M})$ is the likelihood of the data for model $\mathcal{M}$, and C is a constant term. The posterior probability of model $\mathcal{M}$, given sample $\mathbf{s}$ is

$$p(\mathcal{M}|\mathbf{s}) \propto p(\mathbf{s}|\mathcal{M})p(\mathcal{M}).$$

If the prior probability $p(\mathcal{M})$ is the same for any model $\mathcal{M}$, then the model with the highest $p(\mathbf{s}|\mathcal{M})$ is chosen. Suppose we have an unknown parameter $\boldsymbol{\theta}$, then

$$p(\mathbf{s}|\mathcal{M}) = \int_{\boldsymbol{\theta}} p(\mathbf{s}|\boldsymbol{\theta}, \mathcal{M})p(\boldsymbol{\theta}|\mathcal{M})d\boldsymbol{\theta}.$$

The integral is hard to evaluate. BIC is approximated as

$$2l_{\mathcal{M}}(\mathbf{s}, \hat{\boldsymbol{\theta}}) - m_{\mathcal{M}} \log(n), \tag{5.2}$$

where $l_{\mathcal{M}}(\mathbf{s}, \hat{\boldsymbol{\theta}})$ is the maximized log-likelihood of model $\mathcal{M}$; $m_{\mathcal{M}}$ is the number of independent parameters in model $\mathcal{M}$; and n is the sample size. The second term is a penalty of the increased number of parameters. The fit of a model is increased as more parameters are introduced. However, it is not desirable to introduce too many parameters. Thus, we choose the model that maximizes the approximated BIC criteria (5.2). In order to reduce computational effort, the first local maxima

of the criteria is chosen. Figure 11 shows a simulated defects with overlapped local clusters and a graph of the BIC of the data. The first local maxima is obtained at three, the number of expected clusters including the set of global defects. With the BIC, the estimates of the parameters and the number of clusters is obtained simultaneously. The simulation with overlapped local clusters as shown on the left of Figure 11 correctly estimates the number of clusters with 98% accuracy.

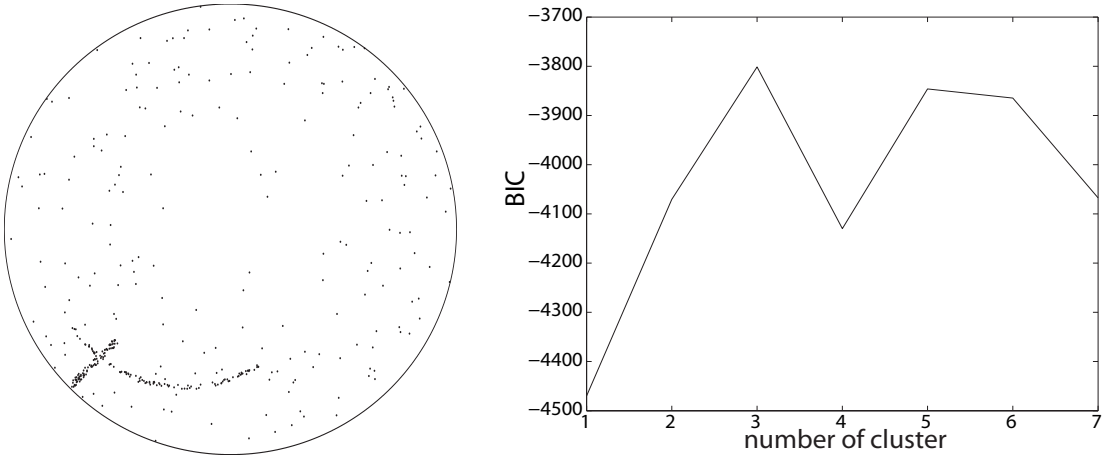


Figure 11: Bayesian information criteria.

V.2.5. Principal curve

The curvilinear feature is one of the typical patterns observed on wafer defect maps with an ellipsoidal pattern. Approximate modeling of curvilinear defect patterns with the bivariate normal distribution may provide incorrect information about the characteristics of clusters and decrease the classification capability [36]. Therefore, in this study, clusters with curvilinear pattern are modeled by the principal curve [49]. The principal curve is a one dimensional curve that passes through d-dimensional data. The one dimensional curve in $\mathcal{R}^d$ is defined as a vector function $\mathbf{f}(\lambda)$ of a scalar

variable λ. Define $\mathbf{s} \in \mathcal{R}^d$, a random vector. Then, $\mathbf{f}(\lambda)$ is the principal curve of $\mathbf{s}$ if

$$E[\mathbf{s}|\lambda_{\mathbf{f}}(\mathbf{s}) = \lambda] = \mathbf{f}(\lambda),$$

where

$$\lambda_{\mathbf{f}}(\mathbf{s}) = \sup_{\lambda}\{\lambda : ||\mathbf{s} - \mathbf{f}(\lambda)|| = \inf_{\mu} ||\mathbf{s} - \mathbf{f}(\mu)||\}.$$

$\lambda_{\mathbf{f}}(\mathbf{s})$ is the projection point λ, i.e., the closest point on $\mathbf{f}(\lambda)$ from $\mathbf{s}$.

The principal curve of data $\mathbf{s}$ of size n is composed of n tuples. The algorithm for finding the principal curve from data $\mathbf{s}$ is [49]:

Initialization:

Find $\mathbf{f}^0(\lambda) = \bar{\mathbf{s}} + \mathbf{a}\lambda$ *where* $\mathbf{a}$ *is the principal component of* X.

Repeat:

Find $\lambda^i_{\mathbf{f}}(\mathbf{s}_j)$, *for* $j = 1, \ldots, n$.

Find $\mathbf{f}^{i+1}(\lambda)$.

Until

$|D^{i+1} - D^i|$ *is less than a criterion where* $D^i = \sum_{j=1}^{n} ||\mathbf{s}_j - \mathbf{f}^i(\lambda^i_{\mathbf{f}}(\mathbf{s}_j))||^2$.

In finding $\mathbf{f}^{i+1}(\lambda)$, the principal curve in the $i+1^{st}$ iteration, the natural way is to get the average of the points that project to the same point on the principal curve. However, with a finite number of data points, we have only one such point. Thus, we use a locally weighted running-line smoother with a fraction ω of points. The weight of $\mathbf{s}_j$ for $\mathbf{s}_i$ is [49]:

$$w_{ij} = \left[1 - \left(\frac{|\lambda_i - \lambda_j|}{h_i}\right)^3\right]^3,$$

where h_i is the distance on the principal curve from λ_i to the ωn^{th} closest point. The

weight is one for the point itself and diminishes as the distance increases. Figure 12 shows the illustration of a principal curve of a two dimensional data.

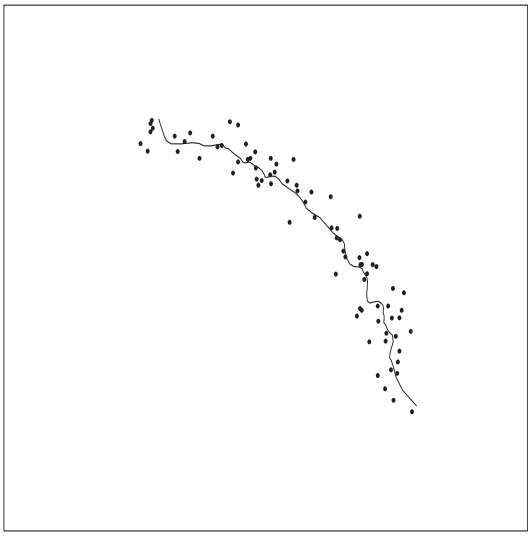


Figure 12: Principal curve.

V.3. The clustering strategy

V.3.1. Global defect pattern and distribution

The global defects take a non-random pattern which has a larger defect density as the distance from the center of the wafer increases. This pattern cannot be modeled as background random noise and, thereby, needs a distribution function that describes its non-random pattern for the model-based clustering. Non-random defects on a wafer can be considered as the realization from a spatial nonhomogeneous Poisson process. In a spatial nonhomogeneous Poisson process, the pattern of event locations is modeled by the intensity function which governs the likelihood of an event occurring at a location. From the intensity function, we get the PDF of the

defect location on the wafer region. Consider a point process $\{N(D) : |D| \geq 0\}$ on a planar region $D \in \mathcal{R}^d$. $N(D)$ is the number of events on D. The intensity at location $\mathbf{s} \in D$, $\lambda(\mathbf{s})$, is defined as [31]

$$\lambda(\mathbf{s}) = \lim_{|d\mathbf{s}|\to 0} \frac{E[N(d\mathbf{s})]}{|d\mathbf{s}|}, \tag{5.3}$$

where $E[N(d\mathbf{s})]$ is the average number of events on a infinitesimal region around $\mathbf{s}$, $d\mathbf{s}$. Then, the PDF of the process is

$$f(\mathbf{s}) = \frac{\lambda(\mathbf{s})}{\int_D \lambda(\mathbf{s}) d\mathbf{s}}. \tag{5.4}$$

The estimators of the parameters, $\boldsymbol{\alpha}$, are obtained by maximizing the likelihood function [33]:

$$L(\boldsymbol{\alpha}|\mathbf{s}) = \prod_{i=1}^{n} \lambda(\mathbf{s}_i) \exp\left\{ - \int_D \lambda(\mathbf{s}) d\mathbf{s} \right\}. \tag{5.5}$$

Generally, a closed form maximum likelihood estimator does not exist.

V.3.2. Local defect patterns

The local defect clusters have specific patterns according to their causes. Particles or stains generate ellipsoidal shape defect patterns [43], and these are well described by a bivariate normal distribution. Linear features can be modeled well by a bivariate normal distribution. However, non-convex objects, like curvilinear patterns, are not modeled well by a bivariate normal distribution. Human induced scratches or bad batch handling cause curvilinear defect patterns [50]. In modeling curvilinear patterns, I use the principal curve that is defined by Hastie and Stuetzle [49]. Applications of clustering with the principal curve can be found in [51, 52]. The principal curve is a generalization of the principal component of a data set. The likelihood of

the principal curve is defined as [52]

$$L(\mathbf{s}|\mathbf{f}) = \prod_{i=1}^{n} L(\mathbf{s}_i|\mathbf{f}),$$

where

$$L(\mathbf{s}_i|\mathbf{f}) = \frac{1}{\nu}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{||\mathbf{s}_i - \lambda_f(\mathbf{s}_i)||^2}{2\sigma^2}\right), \tag{5.6}$$

where ν is the length of $\mathbf{f}$ and $||\mathbf{s} - \lambda_f(\mathbf{s})||$ is the distance from $\mathbf{s}$ to $\mathbf{f}$. It is assumed that the defects are uniformly distributed on $[0, \nu]$ along the principal curve and that they are normally distributed about the principal curve.

V.3.3. Statistical clustering algorithm

The proposed clustering algorithm consists of two steps; the orientation step and the local tuning step. I combine the BIC and EM with a complete-data loglikelihood criterion and a classification loglikelihood criterion in order to develop a stable and refined algorithm. The basic idea is:

- Find the number of clusters and the initial point with BIC using L_1 criterion with the bivariate normal distribution.
- Apply the L_2 criterion with the principal curve distribution, equation (5.6).
- Choose the cluster that yields the larger BIC between the bivariate normal distribution and the principal curve.

In the orientation step, the bivariate normal distribution with a variable covariance matrix is used as the distribution of the cluster in order to find the initial point for the next step. I also apply the BIC to choose the number of clusters. I use the EM algorithm with criterion L_1, equation (5.1). The number of cluster is found by the BIC with the bivariate normal distribution for the local defect clusters. This

step can identify the locations of the clusters but is not suitable for capturing the properties of the clusters with curvilinear patterns. In the local tuning step, the CEM algorithm is applied with the clustering result of the orientation step as the initial point. The discovered clusters are run with the principal curve distribution, which can be obtained from equation (5.6). This step examines whether the cluster feature is closer to a curvilinear pattern than an ellipsoidal pattern. The BIC of each cluster obtained by the principal curve is compared with that of each cluster detected by the bivariate normal distribution in the orientation step to identify the pattern of the cluster. Between the bivariate normal distribution and the principal curve, the model that yields the larger BIC is selected. If the principal curve has a larger BIC, then it is an indication that the cluster has a linear or curvilinear signature. For the convergence criteria for both EM and CEM algorithms, I use the coefficients of global intensity function in equation (5.7).

V.4. Simulated defects

The global defects are generated by the thinning method [37] from random defects on the wafer. I use the quadratic intensity function. The advantage of the quadratic intensity function is the flexibility of the function. A polar coordinate is used with the assumption that the defect pattern is isotropic, i.e., there is no angular term in the intensity function. The intensity function used is

$$\lambda(r) = \alpha_0 + \alpha_1 r + \alpha_2 r^2, \tag{5.7}$$

where r represents the distance from the center of a wafer.

Since I don't have the actual defect data nor a description of the data, I generate local defects that resemble the patterns shown in [50]. These local defects

are generated from distributions that are different from the distributions adopted in the clustering algorithm. The ellipsoidal pattern is generated from an uniform distribution on a rectangle. The curvilinear pattern is generated from the assumption that the defects are distributed uniformly along and about the arc. The double curvilinear pattern is obtained by concatenating a curvilinear pattern and a linear pattern.

Figure 13 shows the the results of clustering with simulated defects. The general pattern of defect clusters is captured by the algorithm, and the parameters of the clusters, such as the center and the covariance for the ellipsoidal pattern and the principal curve for the curvilinear or the amorphous pattern, are obtained. And most importantly, the simulation results identify, even in the worst case, the locations of the clusters.

V.5. Conclusion

This chapter proposes an automated wafer defect data analysis method based on model-based clustering for semiconductor fabrication process control. The model-based clustering with a spatial nonhomogeneous Poisson process, the bivariate normal distribution and the principal curve classify the defects according to their defect generation mechanisms using the spatial patterns of defect clusters. In this procedure, low speed high resolution methods and high speed low resolution methods are combined to maximize the root-cause analysis capability by sorting out those wafers that need further visual inspection and by narrowing down the inspection area on these wafers. Also, the variation in defect densities of the various defect generation mechanisms can be monitored by the parameters of the clusters. The simulation presents a promising result. It successfully finds the number of clusters

and describes the properties of these clusters. Due to the classification capability of this algorithm, complicated defect patterns generated by the superposition of several defect patterns can be modeled with fewer parameters. This result provides useful information about the defect density variations on a wafer, and it is a good basis for yield modeling based on the spatial point process.

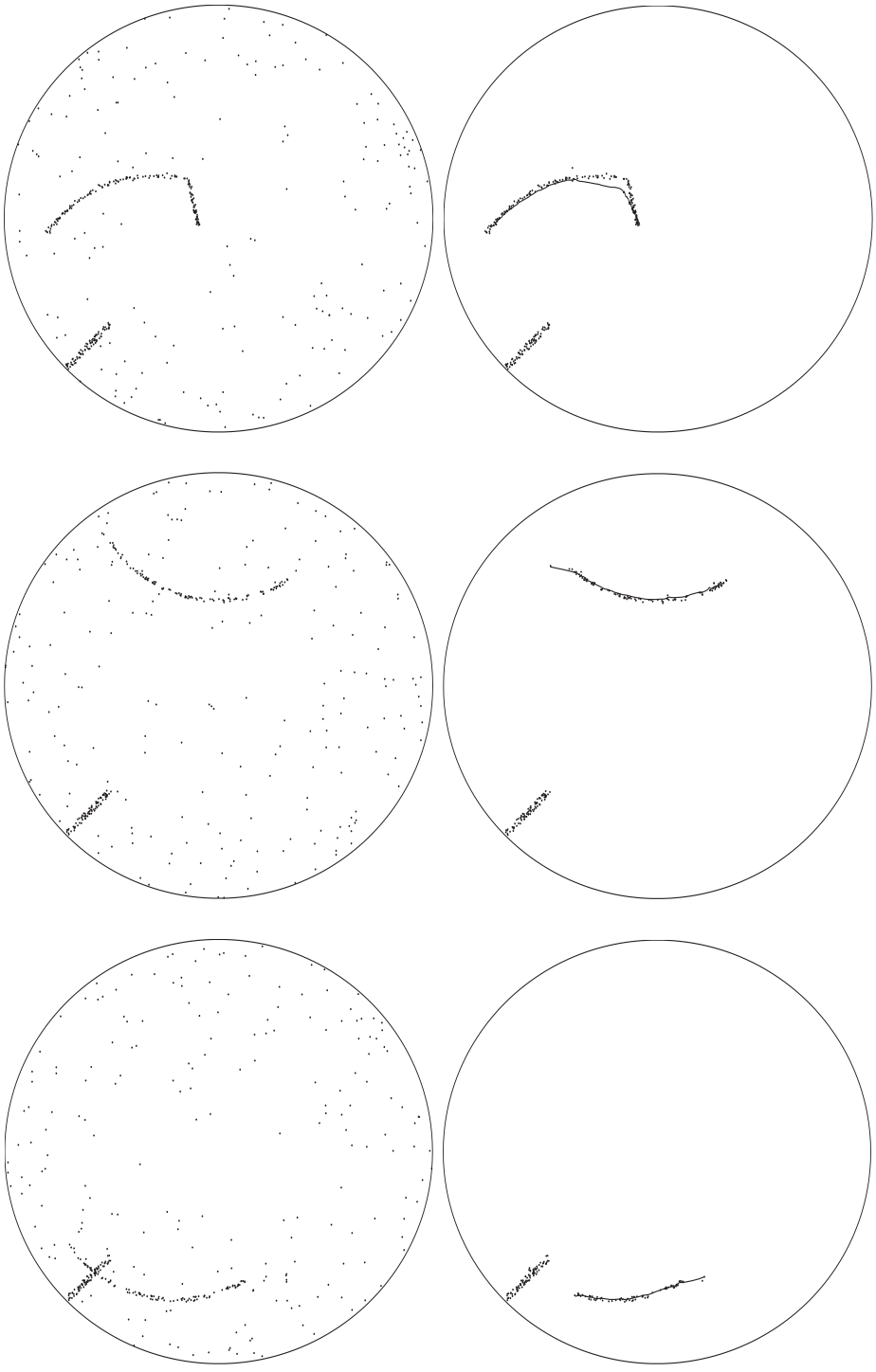


Figure 13: Clustering results of simulated defects.
Figures on the left-hand side are the stacked wafer map, and the figures on the right are the detected clusters with principal curves.

CHAPTER VI

EXTRINSIC RELIABILITY ESTIMATION FROM YIELD INFORMATION

This chapter studies the yield-reliability relation model for estimating the extrinsic reliability of ICs. A yield model based on the spatial nonhomogeneous Poisson process and a statistical defect growth model are introduced. An extrinsic reliability model is derived using the yield model and the statistical defect growth model. The properties of a reliability function, in terms of a failure mechanism using a failure rate function, is investigated.

VI.1. Introduction

Reliability is defined as the probability that a system works properly up to a specified time under specified working condition [5]. Varying reliability by time is mathematically represented by a reliability function. In estimating reliability, statistical inference is preferred to analytical failure models due to the complex failure mechanisms of ICs. Statistical inference for reliability estimation is normally done from life testing data. Parametric methods with distribution functions are widely used. In choosing a parametric distribution family for reliability, one should consider the physical constraints of the failure mechanism, the failure rate function which represents the characteristics of the aging process [21]. An increasing failure rate over time means that the system ages i.e., the system deteriorates in time.

Electronic devices are known to have a bathtub-shaped failure rate curve, and this curve cannot be well modeled by a single distribution function. Some of the literature studies several failure modes with the mixture distribution model [53, 54].

However, the mixture distribution model fails to derive a bathtub-shaped failure rate curve. The competing risk model describes the bathtub-shaped failure rate function using monotone failure rate distributions. With this competing risk model, the failure rate function of a system is obtained with the summation of the failure rate functions of two or three independent failure modes, which have increasing, decreasing or constant failure rate functions. Each failure rate function represents an inherited failure mode in the failure mechanism. An increasing failure rate curve results from the intrinsic failure mode which is an inherited aging property(or wear-out), and a decreasing failure rate curve results from the extrinsic failure mode which causes early failures due to extraneous factors, mostly defects, in the products. A bathtub-shaped failure rate curve means that the extrinsic failure mode dominates the failure mechanisms during the early life of a product.

It is known that electronic devices, including ICs have very long useful life on failure rate curve. Technology in the semiconductor industry has advanced to a level that insures high intrinsic reliability up to the mission time of products [55]. Consequently, estimating reliability using life testing takes a long time, even with accelerated life testing. Considering the long life of well designed and produced IC products, however, only extrinsic reliability is meaningful as feedback to the design and manufacturing process. Extrinsic reliability is determined by the defects introduced during the manufacturing process and has a decreasing failure rate function. Extrinsic reliability and yield have the same source of loss. Thus, the relationship between these two measures and extrinsic reliability estimation from defect data have been studied by several scholars.

Kuper et al. [29] and Van der Pol et al. [28] observe that yield and reliability are positively correlated. An effort to retrieve reliability information from yield have been made by several researchers [24, 25, 26, 27]. This literature can be divided into

two subcategories. One is the model that identifies proportional defect density as the cause of reliability failure in operation. The defect density of latent defects(or reliability defects) is assumed to be proportional to the yield defect density [24, 25]. This assumption consequently restricts the reliability function to be static, i.e., reliability is not a function of time. The other model assumes defect growth under operational conditions [26, 27]. The analytical defect growth model of gate oxide defects has been studied, but it needs modification for other kinds of defects. Reliability critical area models [56, 57] that assume changes of critical areas under stress can also be classified in this subcategory.

The purpose of this study is to build an extrinsic reliability estimation model from yield information. At first, the defect pattern on a wafer is modeled using the spatial nonhomogeneous Poisson process(NHPP). A yield model is constructed based on the spatial NHPP. Then, using a statistical defect growth model, an extrinsic reliability model is derived based on the yield model. Derived extrinsic reliability deals with general defects in an IC and is a function of time. Due to the intensity function that describes the spatial pattern of defects on a wafer, unlike the existing relation models, we can compute a distinct reliability function according to the chip locations on a wafer. Different reliabilities by chip location can be used for screening low reliability wafers and determining adjustable burn-in policies according to the defect density of the specific chip, wafer or lot. Extrinsic reliability estimation doesn't rule out the necessity of life testing. However, this method can provide a more reliable method for reliability estimation.

This chapter is composed as follows: The competing risk model is discussed in section VI.2. Yield modeling based on the spatial nonhomogeneous Poisson process is introduced in section VI.3. Section VI.4 discusses a statistical defect growth model. In section VI.5, reliability estimation from defect data using the new yield model

based on the spatial nonhomogeneous Poisson process is introduced. The failure rate function of extrinsic reliability is examined in section VI.6. Section VI.7 concludes this study and addresses future research.

VI.2. Competing risk model

The property of the aging process of a system can be described by the failure rate curve. The integral of the failure rate function on $[t, t+\Delta t]$ represents the probability that the system in a working state fails in the time interval. Thus, the higher the failure rate, the higher the failure probability. It is known that electronic devices including ICs have bathtub-shaped failure rate curves as shown in Figure 6. This means, in a probabilistic sense, that the population of products rejuvenates as time goes on up to a certain point, retains its age and then deteriorates when it approaches the end of its life span. The decreasing failure rate is caused by defects introduced during the manufacturing process and misuse in operation. The increasing failure rate results from the natural aspects of aging.

The failure mechanism of a system can be explained with the competing risk model. The failure mechanism is subdivided into several independent failure modes. They are considered to be configured in a series system. In other words, the independent failure modes compete to cause the breakdown of a product. Hence, the failure time of a product is the minimum of the failure times of the failure modes. I consider here two failure modes: intrinsic failure and extrinsic failure. Both have monotone failure rates. The extrinsic failure mode has a decreasing failure rate. This failure mode is caused mostly by defects and dominates the failure rate in the early life of a product. The intrinsic failure mode has an increasing failure rate and dominates wear-out. Each failure mode may contain several failure mechanisms, and

there may exist interactions between these failure mechanisms. However, they are treated as a single failure mechanisms. The notation for the derivation of the failure rate function using the competing risk model is defined as follows:

- $T_E,\ F_E(t)$: a random variable, the failure time caused by the extrinsic failure mode and its cumulative distribution function(CDF)
- $T_I,\ F_I(t)$: a random variable, the failure time caused by the intrinsic failure mode and its CDF
- $T,\ F(t)$: a random variable, the failure time $\equiv \min\{T_I, T_E\}$ and its CDF
- $R(\cdot) \equiv 1 - F(\cdot)$
- $h_j(t),\ h(t)$: the failure rates of $T_j,\ j = E, I$ and T, respectively

From the assumption of the failure time of the system, we can get the reliability function:

$$\begin{aligned} F(t) &= 1 - \Pr\{T > t\} \\ &= 1 - \Pr\{\min\{T_I, T_E\} > t\} \\ &= 1 - \Pr\{T_i > t \text{ and } T_e > t\} \\ &= 1 - \Pr\{T_i > t\} \Pr\{T_e > t\} \\ &= 1 - R_I(t)R_E(t). \end{aligned} \tag{6.1}$$

Assuming $F_j(t)$ has the probability density function(PDF), $\frac{dF_j(t)}{dt} = f_j(t),\ j = I, E,$

$$f(t) = f_I(t)R_E(t) + f_E(t)R_I(t). \tag{6.2}$$

We can get the failure rate of T from equation (6.1) and (6.2):

$$h(t) = \frac{f(t)}{R(t)} = h_I(t) + h_E(t). \tag{6.3}$$

We have a failure rate which is a summation of the two failure rate functions as expected.

VI.3. Yield model based on spatial NHPP

VI.3.1. Spatial nonhomogeneous Poisson process

In order to estimate yield, three things must be known: the spatial distribution of the defect density; the defect size distribution; and the interaction of the above information with the layout of the IC. These are the factors of compound Poisson yield models as seen in previous sections. The defect density distribution which is not well modeled by compound Poisson yield models can be nicely described by spatial point processes. In new yield models based on spatial NHPP, defect clustering is explained by the higher defect density areas that produce aggregated defects. Useful concepts and definitions of the spatial point process are summarized below from Diggle [32] and Cressie [31].

The spatial process is used to model spatial data. The general spatial process model is

$$\{Z(\mathbf{s}) : \mathbf{s} \in D\},$$

where $Z(\mathbf{s})$ is the random quantity at a location $\mathbf{s} \in D$. In many cases, among many possible measures of the events, the locations of events $\mathbf{s}_1, \mathbf{s}_2, \ldots$ on $D \subset \mathcal{R}^d$ are the main concern. The locations can be modeled by the spatial point process as follows: Let $(\Omega, \mathcal{F}, P)$ be a probability space and let Φ be a collection of locally finite

counting measures on $D \subset \mathcal{R}^d$. On Φ, define $\mathcal{N}$, the smallest σ-algebra generated by sets of the form $\{\phi \in \Phi : \phi(B) = n\}$ $\forall B \subset D$ and $\forall n \in \{0, 1, 2, 3, \ldots\}$. Then, a spatial point process on D is a measurable mapping of $(\Omega, \mathcal{F})$ into $(\Phi, \mathcal{N})$. If an event located at $\mathbf{s} \in \mathcal{R}^d$ is marked by $x \in \mathcal{A}$, then $(\mathbf{s}, x)$ is a point in $(\mathcal{R}^d, \mathcal{A})$. So a marked point process is a point process on the product space $(\mathcal{R}^d, \mathcal{A})$.

Consider a point process $\{N(B) : |B| \geq 0\}$ on a planar region $B \in \mathcal{R}^2$. $N(B)$ is the number of events on B. The events of interest in this model are the locations of the defect on the wafer. The size of the defects is dealt with by the marked point process. A NHPP $N(B)$ is defined on a finite planar region B with the following postulates:

1. $N(B)$ has a Poisson distribution with mean $\mu(B) = \int_B \lambda(\mathbf{s}) d\mathbf{s}$, where $\mathbf{s} \in B$ and $\lambda(\mathbf{s})$ is the intensity function on B defined as equation (6.4).

2. Given $N(B) = n$, the n events form an independent random sample from the distribution on B with a PDF proportional to $\lambda(\mathbf{s})$.

The intensity function $\lambda(\mathbf{s})$ is defined as

$$\lambda(\mathbf{s}) \equiv \lim_{|d\mathbf{s}| \to 0} \frac{E[N(d\mathbf{s})]}{|d\mathbf{s}|}. \tag{6.4}$$

Thus, we can compute the average number of defects on B with

$$E[N(B)] = \int_B \lambda(\mathbf{s}) d\mathbf{s}.$$

VI.3.2. Yield model

Let's consider a yield model based on the results shown above. In compound Poisson yield models, the defect variation is modeled by the defect density distribution which

doesn't have spatial information about the defect density variation. The new yield models basically use the Poisson yield model. However, the average number of faults varies die by die, and it is computed using the intensity function of the spatial NHPP. Since the average number of faults is not identical for all dies, we need to modify the definition of yield as the ratio of the number of conforming products at the end of a production process to the number of potentially usable products [7].

The following is the mathematical derivation of the yield model: Define $\{B_j\}_1^k$ as the set of finite disjoint regions of the die area on the wafer, where k is the number of dies on a wafer. Let's define a random variable

$$X_{B_j} = \begin{cases} 1 & \text{if } B_j \text{ is good,} \\ 0 & \text{if } B_j \text{ is defective, } j = 1, \cdots, k. \end{cases}$$

Then, the chip yield of die area B_j given $\lambda(\mathbf{s})$ is

$$\begin{aligned} E[X_{B_j}] &= \Pr\{X_{B_j} = 1\} \\ &= \Pr\{\text{no faults on die } B_j\} \\ &= e^{-\mu(B_j)}, \quad j = 1, \cdots, k, \end{aligned} \tag{6.5}$$

where $\mu(B_j)$ is the parameter of the Poisson distribution and means the average number of faults on die B_j. The equation (6.5) holds by the first postulate of the spatial NHPP.

To build a yield model, let's define a new random variable $X = \sum_{i=1}^{k} X_{B_i}$; then X represents the number of conforming dies on a wafer. Thus, we have a yield

model defined as the average portion of conforming dies per wafer:

$$\begin{aligned} Y &= \frac{E[X]}{k} \\ &= \frac{E[X_{B_1}] + E[X_{B_2}] + \cdots + E[X_{B_k}]}{k}. \end{aligned} \tag{6.6}$$

The following is a method for getting $\mu(B)$, the average number of faults on an arbitrary chip region B. Given the location of a defect, a bigger defect is more likely to cause a circuit failure. In order to deal with various defect sizes, the marked point process is adopted. The size of the defect x is the marked random quantity of the marked point process. The intensity of the marked point process on the product space $(\mathcal{R}^2 \times \mathcal{A})$ is defined as

$$\nu(\mathbf{s}, x) \equiv \lim_{|d\mathbf{s} \times dx| \to 0} \frac{E[N(d\mathbf{s} \times dx)]}{|d\mathbf{s} \times dx|}.$$

Then $\nu(\mathbf{s}, x) d\mathbf{s} dx$ is the average number of defects of size x on region $\mathbf{s}$, and the average number of faults of size x on region $\mathbf{s}$ is

$$\frac{A_c(x)}{|B|} \nu(\mathbf{s}, x) d\mathbf{s} dx.$$

Consequently, the average number of faults on region B, $\mu(B)$ is

$$\mu(B) = \int_0^\infty \int_B \frac{A_c(x)}{|B|} \nu(\mathbf{s}, x) d\mathbf{s} dx. \tag{6.7}$$

Note that the intensity function of the marked Poisson process is factorized as

$$\nu(\mathbf{s}, x) = \lambda(\mathbf{s}) g(x|\mathbf{s}),$$

where $g(x|\mathbf{s})$ is the conditional distribution of the defect size given the defect location.

If $g(x|\mathbf{s}) = g(x)$, then

$$\mu(B) = \int_0^\infty A_c(x)g(x)dx\frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|},$$

and the chip yield is obtained as

$$E[X_B] = \exp\left\{-\int_0^\infty A_c(x)g(x)dx\frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|}\right\}.$$

Now, the yield can be computed from equation (6.6) using the chip yields.

VI.4. Defect growth model

VI.4.1. Mixture defect size distribution

The extrinsic reliability model is based on the severity of defects modeled by the defect growth model. The behavior of defect growth is modeled by the defect size distribution at time t, $g_t(x)$, the PDF of X_t, and the random variable of defect size at time t. Analytical defect growth models have been applied to gate oxide reliability [26, 27]. In this study, all kinds of defects are considered. Some kinds of defects don't change in severity(size) under operational stress [25], and the defect growth rate of different kinds of defects may not necessarily be the same. Also, the composition of defects may vary. Thus, in order to incorporate several defects that have different growth rates, a statistical defect growth model which is obtained by a mixture distribution of the defect groups with different defect growth rates is considered.

I categorize the defects into "growing defects" and "dormant defects." The defect size of a dormant defect doesn't change in time under operation. Let X_t^*, X_t^{**} be the random variables of the defect size of a growing defect and a dormant defect,

respectively, and $g_t^*(x), g_t^{**}(x)$ be the corresponding PDF's at time t. Denote k as the proportion of growing defects to the whole number of defects. Then, we have

$$g_t(x) = kg_t^*(x) + (1-k)g_t^{**}(x). \tag{6.8}$$

It is assumed that the defect size distribution at $t = 0$ is known.

VI.4.2. Statistical defect growth model

Since the defect size distribution of dormant defects stays the same in time, i.e., $g_t^{**}(x) = g_0^{**}(x)$ for any $t > 0$ and $x \geq 0$, $g_t(x)$ is obtained by equation (6.8) with the transformation of $g_0^*(x)$ with $X_t^* = h(X_0^*)$. The function $h(\cdot)$ is a function of X_0^*, time and stress level, and the time and stress level are modeled through the defect growth factor, $\eta(t, e)$ at time t under stress level e. There exists no well-known defect growth model for all kinds of defects. However, it can be assumed that the defect growth mechanism is analogous to that of gate oxide defects [58]

$$t_{BD} = \tau \exp\left\{\frac{G}{V}(w-x)\right\}, \tag{6.9}$$

where t_{BD} is the breakdown time; G is a constant with a unit of voltage/length; V is the voltage applied to the oxide; w is the thickness of the oxide; x is the defect size; and τ is a constant which has a unit in time. Then the breakdown time, t_{BD}, can be interpreted as the time elapsed for the defect growth of $w - x$, and we get the defect growth model:

$$X_t^* = X_0^* + \eta(t, e),$$

with the relation:

$$\eta(t, e) \propto e \ln t,$$

where e in this model is the voltage.

VI.5. Extrinsic reliability model

In this section, the extrinsic reliability model is derived. Yield and extrinsic reliability have the same source of loss-defects. Reliability is a function of the operation time. If yield and reliability are put on the same line of time flow, yield can be viewed as reliability at $t = 0$. However, it is worth noting that the starting time, $t = 0$, when the system is put into operation, is considered in a relative sense. That is, the starting time is not fixed, and it can be adjusted by electrical tests and burn-in.

It is assumed that the intensity function of spatial NHPP, $\lambda(\mathbf{s})$, doesn't change in time, i.e., we don't observe new defects during operation. Although we may have new defects in a new location, the probability that the new defects break down the chip earlier than the existing defects is assumed to be negligible. Suppose that $t = 0$ represents the time when the products are produced and the test is over. Then, the reliability of a chip on region B at time t is defined using conditional reliability as

$$\begin{aligned} R(t) &\equiv \Pr\{T > t | T > 0\} \\ &= \frac{\Pr\{T > t\}}{\Pr\{T > 0\}}, \text{ if } \Pr\{T > 0\} > 0. \end{aligned} \tag{6.10}$$

We can compute $\Pr\{T > t\}$ using a Poisson distribution with the average number of faults on chip B at time t, $\mu(B, t)$:

$$\begin{aligned} \Pr\{T > t\} &= \Pr\{\text{the chip works without failure up to } t\} \\ &= \Pr\{\text{no faults on chip up to } t\} \\ &= e^{-\mu(B,t)}. \end{aligned}$$

$\mu(B,t)$ is computed from equation (6.7):

$$\mu(B,t) = A_c(t)\frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|},$$

where $A_c(t)$ is the average critical area of region B at time t computed as

$$A_c(t) = \int_0^\infty A_c(x)g_t(x)dx.$$

We have

$$\begin{aligned}\Pr\{T > t\} &= \exp\left\{-A_c(t)\frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|}\right\} \\ &= \exp\left\{-\int_0^\infty A_c(x)g_t(x)dx\,\frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|}\right\}, \qquad (6.11)\end{aligned}$$

and

$$\begin{aligned}\Pr\{T > 0\} &= E[X_B] \\ &= \exp\left\{-A_c(0)\frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|}\right\} \\ &= \exp\left\{-\int_0^\infty A_c(x)g_0(x)dx\,\frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|}\right\}. \qquad (6.12)\end{aligned}$$

Therefore, the reliability function obtained from equations (6.10),(6.11) and (6.12) is

$$\begin{aligned}R(t) &= \frac{\exp\left\{-\int_0^\infty A_c(x)g_t(x)dx\,\frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|}\right\}}{\exp\left\{-\int_0^\infty A_c(x)g_0(x)dx\,\frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|}\right\}} \\ &= \exp\left\{-\int_0^\infty A_c(x)[g_t(x) - g_0(x)]dx\,\frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|}\right\} \\ &= \exp\left\{-\int_0^\infty A_c(x)[g_t^*(x) - g_0^*(x)]dx\,\frac{k\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|}\right\}. \qquad (6.13)\end{aligned}$$

The reliability function is inversely proportional to the defect density on the chip and

the ratio of growing defects. Also, it is inversely proportional to the defect growth rate $\eta(t, e)$, which is not explicitly shown on the reliability function.

The relationship between yield and reliability at a fixed time with various k values is illustrated in Figure 14. The shape of the relation curve varies by k. Given the same yield, the smaller the k value, the higher the reliability. We also observe that the extrinsic reliability is bounded from below. With the assumption that $A_c(t) \to |B|$ as $t \to \infty$, the extrinsic reliability of the chip on B is bounded from below, and as $t \to \infty$

$$R(t) \to \frac{\exp\{-\int_B \lambda(\mathbf{s})d\mathbf{s}\}}{\exp\left\{-A_c(0)\frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|}\right\}}.$$

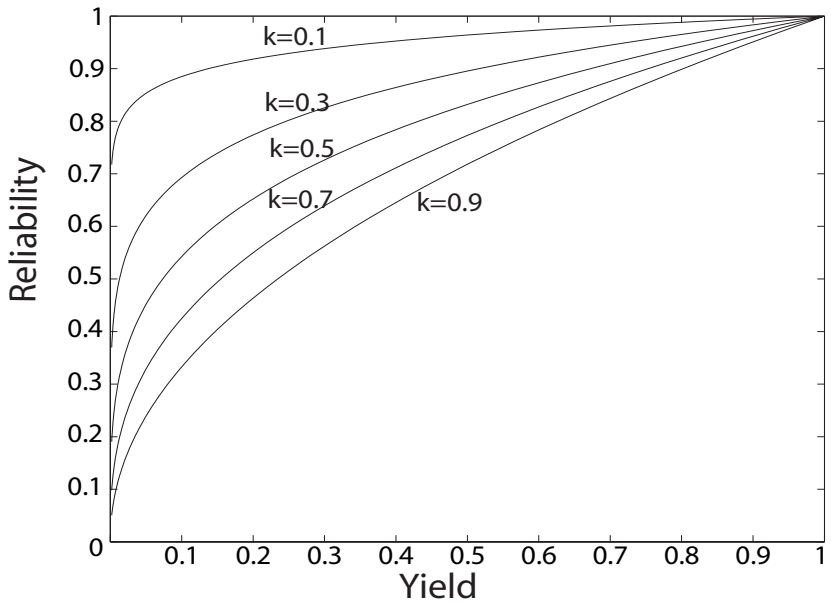


Figure 14: Yield and reliability relation with various k values.

VI.6. Failure rate function

Extrinsic reliability has a decreasing failure rate. The reliability derived from the defect data should be in the decreasing failure rate class. Since $f(t) = -\dfrac{dR(t)}{dt}$ when the derivative exits,

$$\begin{aligned} f(t) &= -\frac{d \exp\left\{-\int_0^\infty A_c(x)[g_t(x) - g_0(x)]dx \, \frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|}\right\}}{dt} \\ &= R(t)\frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|} \; \frac{d\int_0^\infty A_c(x)g_t(x)dx}{dt}. \end{aligned}$$

Thus, the failure rate function is

$$\begin{aligned} h(t) &= \frac{f(t)}{R(t)} \\ &= \frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|} \; \frac{d\int_0^\infty A_c(x)g_t(x)dx}{dt} \\ &= \frac{\int_B \lambda(\mathbf{s})d\mathbf{s}}{|B|} \; \frac{dA_c(t)}{dt}. \end{aligned}$$

If the average critical area at t, $A_c(t)$ is concave, $h(t)$ is decreasing in t. Since the closed form of the failure rate function is not available, the property of the extrinsic failure rate curve is examined numerically with the following settings.

The defect size distribution is assumed as [11]

$$g(x) = \begin{cases} \dfrac{2(p-1)x}{(p+1)x_0^2}, & \text{if } x \le x_0, \\ \dfrac{2(p-1)x_0^{p-1}}{(p+1)x^p}, & \text{if } x > x_0. \end{cases}$$

Let $p = 3$ and $\eta(t, e) = \dfrac{e}{a} \log \dfrac{t}{b}$. Then

$$g_t^*(x) = \begin{cases} \dfrac{x - \eta(t,e)}{\tilde{x}_0^2}, & \text{if } \eta(t,e) \le x \le \tilde{x}_0, \\ \dfrac{\tilde{x}_0^2}{(x - \eta(t,e))^3}, & \text{if } x > \tilde{x}_0, \end{cases}$$

where $\tilde{x}_0 = x_0 + \dfrac{e}{a} \log \dfrac{t}{b}$. And the critical area model is assumed as

$$A_c(x) = \begin{cases} 0, & \text{if } x < w_0, \\ \dfrac{x - w_0}{w_1 - w_0}|B|, & \text{if } w_0 \le x < w_1, \\ |B|, & \text{if } x \ge w_1, \end{cases}$$

with $|B| = 1$. I assume that the critical area is linear to the defect size x such that $w_0 \le x \le w_1$ and is bounded by the die area. Figure 15 shows the failure rate curve of the extrinsic reliability derived numerically. It shows the monotone decreasing

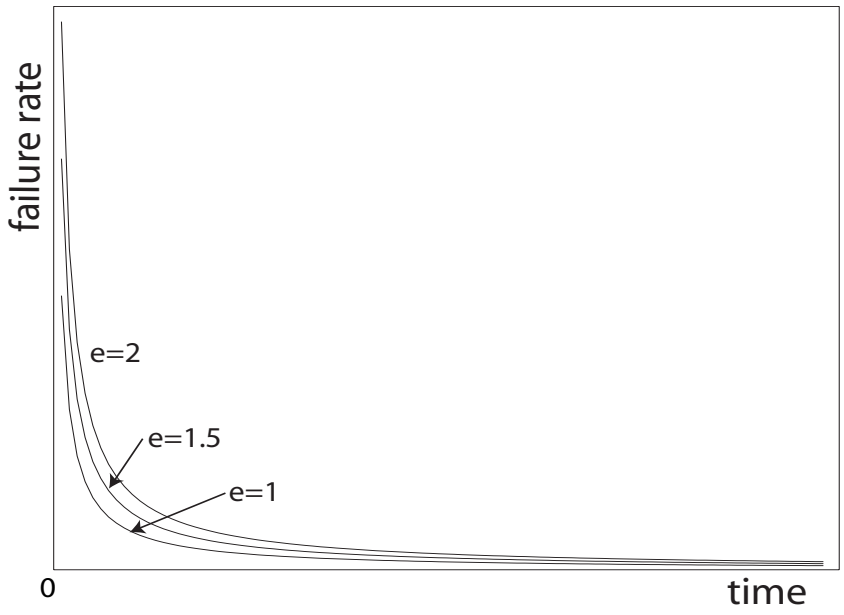


Figure 15: Extrinsic failure rate curve in arbitrary units with $\eta(t, e) = \dfrac{e}{a} \log \dfrac{t}{b}$.

property as desired.

VI.7. Conclusion and future research

The relationship between yield and reliability based on defect data using the statistical defect growth model has been examined here. Extrinsic reliability was derived from the new yield model based on the spatial NHPP. With the extrinsic reliability model based on the spatial NHPP, the positive correlation between yield and reliability was explained. Since this model utilizes the additional information of defect data, it contributes to more accurate life testing at a reduced cost. Estimating reliability, especially extrinsic reliability, which is closely related to most failures in the early life of ICs, provides important feedback to circuit design and the determination of optimal burn-in times. Without time consuming life testing of ICs, this method can shorten the feedback loop. More importantly, in this model, the extrinsic reliability of a specific chip depends on the defect density on the chip location. Thus, based on the yield of the wafer or the lot, different screening rules for low reliability chips and for burn-in policy can be applied.

The following directions are suggested for future research. In this study, it is assumed that the intensity function doesn't change under operational conditions. This assumption is based on the underlying assumption that current inspection technology is powerful enough to detect small defects that may cause circuit failures in operation. However, due to the scaling of ICs, it is likely in the near future that undetectably small defects may cause circuit failures [59]. In this case, the intensity function should be modified in time to model the defects that actually exist on a wafer but are not counted in the yield modeling.

CHAPTER VII

CONCLUDING REMARKS

This dissertation studies the variation management methods of nanotechnology using information about the spatial defect patterns, and the methods are applied in semiconductor manufacturing to yield modeling, defect pattern recognition and extrinsic reliability estimation. So far, the defect density variation in yield models is modeled in compound distributions which simply represent the variation in defect density without considering the spatial defect patterns within and between wafers. Defect pattern recognition has been studied for IC fabrication process control, especially for root-cause analysis, independently of yield modeling. This study develops a method of modeling spatial defect patterns for an unified scheme of process variation management that includes yield and reliability modeling and root-cause analysis.

The spatial defect pattern is modeled using the spatial point process. Considering the nonhomogeneity of the defect density on wafers, the spatial nonhomogeneous Poisson process is used. The intensity function, which is a function of the location of the nonhomogeneous Poisson process, governs the spatial distribution of defects. And in order to incorporate the various defect sizes on a wafer and the various intensity functions between wafers, a spatial marked point process and a doubly stochastic process are adopted. The defect density variation by location is modeled by the intensity function of the process, and the clustering effect can be explained by the regions of higher intensity functions. The yield model, pattern recognition algorithm and extrinsic reliability estimation method are based on the intensity function of the spatial nonhomogeneous Poisson process.

Our new yield model, based on the spatial nonhomogeneous Poisson process,

captures the defect density variation using the intensity function which is estimated from a maximum likelihood estimation with a model-based clustering algorithm. The parameters of the intensity function estimated using model-based clustering contain useful information on the sources of yield loss. Since the intensity function contains spatial information about the defect patterns, the intensity functions estimated in different process steps or in different kinds of defects can be added to get the overall intensity function without worrying about different clustering effects. Unlike existing yield models, each chip has a different yield according to the intensity function evaluation of the chip, and the yield is defined as the average of chip yields on a wafer.

The model-based clustering algorithm also adopted in the intensity function estimation is used in pattern recognition for finding the assignable causes of yield loss. This algorithm facilitates the root-cause analysis by identifying wafers or regions on wafers that need further visual inspection. The defects generated by the local defect generation mechanisms, which are normally the assignable causes, are classified as the defect pattern on the wafer. The characteristics of the defect clusters are represented by the parameters of a bivariate normal distribution or principal curve.

Extrinsic reliability is estimated from the defect data using the statistical defect growth model. The defect growth rate of different defects are considered statistically using a transformation of the defect size distribution. Using the extrinsic reliability model, a positive correlation between yield and reliability can be explained. The benefits of the extrinsic reliability estimation model result from the high intrinsic reliability during the useful life of the IC products. Extrinsic reliability has an important influence on the feedback to circuit design and the burn-in decisions. This model enables reliability estimation from yield data without expensive life testing. And under special assumptions of intrinsic reliability, optimal burn-in time can also

be determined from extrinsic reliability. Since the reliability depends on the intensity function of the defects, different policies of warranty and burn-in can be applied based on the defect density of the chips.

The contributions of this dissertation are (1) providing better process monitoring measures that lead to data reduction of wafer map data (2) combining methods of yield modeling and root-cause analysis that share key information (3) facilitating yield learning processes by providing a fast and rigorous defect classification method based on model-based clustering and (4) providing an extrinsic reliability estimation method that can provide fast feedback to integrated circuit design in the early developmental stages. In summary, this study furthers the understanding of process variation using spatial information about defects and provides a unified framework for defect management.

REFERENCES

[1] National Research Council, *Implications of Emerging Micro- and Nanotechnologies*. Washington, D.C.: National Academies Press, 2002.

[2] L.F. Atherton and R.W. Atherton, *Wafer Fabrication: Factory Performance and Analysis*. Boston: Kluwer Academic Publishers, 1995.

[3] C. Weber, "Yield learning and the sources of profitability in semiconductor manufacturing and process development," in *Proc. IEEE International Symposium on Semiconductor Manufacturing Conference*, pp. 324–329, 2002.

[4] S.L. Riley, "Optical inspection of wafers using large-area defect detection and sampling," in *International Workshop on Defect and Tolerance in VLSI systems*, pp. 12–21, 1992.

[5] W. Kuo, W.K. Chien, and T. Kim, *Reliability, Yield, and Stress Burn-in*. Boston: Kluwer Academic Publishers, 1998.

[6] J. A. Cunningham, "The use and evaluation of yield models in integrated circuits manufacturing," *IEEE Trans. Semiconductor Manufacturing*, vol. 3, May 1990, pp. 60–71.

[7] W. Kuo and T. Kim, "An overview of manufacturing yield and reliability modeling for semiconductor products," *Proc. IEEE*, vol. 87, Aug. 1999, pp. 1329–1344.

[8] M. Raghavachari, A. Srinivasan, and P. Sullo, "Poisson mixture yield models for integrated circuits: a critical review," *Microelectronics and Reliability*, vol. 37, no. 4, 1997, pp. 565–580.

[9] C. K. Hansen and P. Thyregod, "Modeling and estimation of wafer yields and defect densities from microelectronics test structure data," *Quality and Reliability Engineering International*, vol. 12, no. 1, 1996, pp. 9–17.

[10] A. V. Ferris-Prabhu, "Modeling the critical area in yield forecasts," *IEEE Journal of Solid State Circuits*, Aug. 1985, pp. 874–878.

[11] C.H. Stapper, "Modeling of defects in integrated circuit photolithographic patterns," *IBM Journal of Research and Development*, vol. 28, Jul. 1984, pp. 461–475.

[12] A. V. Ferris-Prabhu, "Defect size variations and their effect on the critical area of VLSI devices," *IEEE Journal of Solid State Circuits*, Aug. 1985, pp. 878–880.

[13] A.V. Ferris-Prabhu, *Introduction to Semiconductor Device Yield Modeling*. Boston: Artech House, 1992.

[14] B.T. Murphy, "Cost-size optima of monolithic integrated circuits," *Proc. IEEE*, vol. 52, 1964, pp. 1537–1543.

[15] R.B. Seeds, "Yield, economic and logistic models for complex digital arrays," *IEEE International Convention Record*, vol. 6, 1967, pp. 61–66.

[16] T.L. Michalka, R.C. Varshney, and J.D. Meindl, "A discussion of yield modeling with defect clustering, circuit repair, and circuit redundancy," *IEEE Trans. Semiconductor Manufacturing*, vol. 3, no. 3, 1990, pp. 116–127.

[17] C.H. Stapper, "On Murphy's yield integral," *IEEE Trans. Semiconductor Manufacturing*, vol. 4, Nov. 1991, pp. 294–297.

[18] C. N. Berglund, "A unified yield model incorporating both defect and parametric effects," *IEEE Trans. Semiconductor Manufacturing*, vol. 9, no. 3, 1996, pp. 447–454.

[19] P.R. Pukite and C.L. Berman, "Defect cluster analysis for wafer-scale integration," *IEEE Trans. Semiconductor Manufacturing*, vol. 3, no. 3, 1990, pp. 128–135.

[20] M. H. Hansen, V. N. Nair, and F. J. Friedman, "Monitoring wafer map data from integrated circuit fabrication processes for spatially clustered defects," *Technometrics*, vol. 39, Aug. 1997, pp. 241–253.

[21] R.E. Barlow and F. Proschan, *Mathematical Theory of Reliability.* Philadelphia: SIAM, 1965.

[22] R.E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing.* New York: Holt, Rinehart and Winston, 1975.

[23] F. Guess, E. Walker, and D. Gallant, "Burn-in to improve which measure of reliability?," *Microelectronics Reliability*, vol. 32, no. 6, 1992, pp. 759–762.

[24] T. Barnett, A.D. Singh, and V. Nelson, "Extending integrated-circuit yield-models to estimate early-life reliability," *IEEE Trans. Reliability*, vol. 52, Sep. 2003, pp. 296–300.

[25] H.H. Huston and C.P. Clarke, "Reliability defect detection and screening during processing-theory and implementation," in *Annual Proceedings of Reliability Physics*, pp. 268–275, 1992.

[26] K.O. Kim, W. Kuo, and W. Luo, "A relation model of gate oxide yield and reliability," *Microelectronics Reliability*, vol. 44, 2004, pp. 425–434.

[27] T. Kim, W. Kuo, and W. K. Chien, "A relation model of yield and reliability for the gate oxide failures," in *Proc. Annual Reliability and Maintainability Symposium*, pp. 428–433, 1998.

[28] J.A. van der Pol, F.G. Kuper, and E.R. Ooms, "Relation between yield and reliability of integrated circuits and application to failure rate assessment and reduction in the one digit FIT and PPM reliability era," *Microelectronics Reliability*, vol. 36, no. 11, 1996, pp. 1603–1610.

[29] F. Kuper, J. Van der Pol, E. Ooms, T. Johnson, R. Wijburg, W. Koster, and D. Johnston, "Relation between yield and reliability of integrated circuits: experimental results and application to continuous early failure rate reduction programs," in *Proc. Int. Reliability Physics Symp.*, pp. 17–21, 1996.

[30] T. Kim and W. Kuo, "Modeling manufacturing yield and reliability," *IEEE Trans. Semiconductor Manufacturing*, vol. 12, no. 4, 1999, pp. 485–492.

[31] Noel A.C. Cressie, *Statistics for Spatial Data.* New York: John Wiley & Sons, 1993.

[32] P. J. Diggle, *Statistical Analysis of Spatial Point Patterns.* New York: Academic Press, 1983.

[33] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes.* New York: Springer, 2003.

[34] J.G. Campbell, C. Fraley, F. Murtagh, and A.E. Raftery, "Linear flaw defection in woven textiles using model-based clustering," *Pattern Recognition Letters*, vol. 18, 1997, pp. 1539–1548.

[35] A. Dasgupta and A.E. Raftery, "Detecting features in spatial point processes with clutter via model-based clustering," *Journal of American Statistical Association*, vol. 93, no. 441, 1998, pp. 294–302.

[36] C. Fraley and A.E. Raftery, "Model-based clustering, discriminant analysis, and density estimation," Tech. Rep. 380, Department of Statistics, University of Washington, 2000.

[37] P.A.W. Lewis and G.S. Sheldler, "Simulation of nonhomogeneous Poisson processes by thinning," *Naval Research Logistics Quarterly*, vol. 26, 1979, pp. 403–413.

[38] J.D. Banfield and A.E. Raftery, "Model-based Gaussian and non-Gaussian clustering," *Biometrics*, vol. 49, Sept. 1993, pp. 803–821.

[39] G. Celeux and G. Govaert, "Gaussian parsimonious clustering models," *Pattern Recognition*, vol. 28, no. 5, 1995, pp. 781–793.

[40] M. Baron, C.K. Lakshminarayan, and Z. Chen, "Markov random fields in pattern recognition for semiconductor manufacturing," *Technometrics*, vol. 43, Feb. 2001, pp. 66–72.

[41] F.L. Chen and S.F. Lin, "A neural-network approach to recognize defect spatial pattern in semiconductor fabrication," *IEEE Trans. Semiconductor Manufacturing*, vol. 13, Aug. 2000, pp. 366–373.

[42] S. Cunningham and S. MacKinnon, "Statistical methods for visual defect metrology," *IEEE Trans. Semiconductor Manufacturing*, vol. 11, Feb. 1998, pp. 48–53.

[43] S.S. Gleason, K.W. Tobin, T.P. Karnowski, and F. Lakhani, "Rapid yield learning through optical defect and electrical test analysis," in *Proc. SPIE - The International Society for Optical Engineering*, vol. 3332, pp. 232–242, 1998.

[44] C. K. Hansen and P. Thyregod, "Use of wafer maps in integrated circuit manufacturing," *Microelectronics Reliability*, vol. 38, 1998, pp. 1155–1164.

[45] W. Taam and M. Hamada, "Detecting spatial effects from factorial experiments: An application from integrated circuit manufacturing," *Technometrics*, vol. 35, May 1993, pp. 149–160.

[46] C. Fraley and A.E. Raftery, "How many clusters? Which clustering method? Answers via model-based cluster analysis," Tech. Rep. 329, Department of Statistics, University of Washington, 1998.

[47] G. Celeux and G. Govaert, "A classification EM algorithm for clustering and two stochastic versions," *Computational Statistics & Data Analysis*, vol. 14, 1992, pp. 315–332.

[48] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, Mar. 1978, pp. 461–464.

[49] T. Hastie and W. Stuetzle, "Principal curves," *Journal of American Statistical Association*, vol. 84, Jun. 1989, pp. 502–516.

[50] K.W. Tobin, S.S. Gleason, T.P. Karnowski, and D. Guidry, "Using SSA to measure the efficacy of automated defect data gathering," *Micro*, Apr. 1998.

[51] J.D. Banfield and A.E. Raftery, "Ice floe identification in satellite images using mathematical morphology and clustering about principal curves," *Journal of American Statistical Association*, vol. 87, Mar. 1992, pp. 7–16.

[52] D. Stanford and A.E. Raftery, "Finding curvilinear features in spatial point patterns: Principal curve clustering with noise," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, Jun. 2000, pp. 601–609.

[53] C.M. Kim and D.S. Bai, "Analyses of accelerated life test data under two failure modes," *International Journal of Reliability, Quality and Safety Engineering*, vol. 9, no. 2, 2002, pp. 111–125.

[54] S.S. Menon and K.F. Poole, "Early-life reliability prediction methodology for integrated circuits," *Microelectronics Reliability*, vol. 35, no. 8, 1995, pp. 1147–1155.

[55] J.A. Van der Pol, E.R. Ooms, T. Van't Hof, and F.G. Kuper, "Impact of screening of latent defects at electrical test on the yield-reliability relation and application to burn-in elimination," in *Proc. Int. Reliability Physics Symp.*, pp. 370–377, 1998.

[56] E. Bruls, "Quality and reliability impact of defect data analysis," *IEEE Trans. Semiconductor Manufacturing*, vol. 8, May 1995, pp. 121–129.

[57] A.A. Gerard and A.J. Walton, "Critical area extraction for soft fault estimation," *IEEE Trans. Semiconductor Manufacturing*, vol. 11, Feb. 1998, pp. 146–154.

[58] J.C. Lee and I.C. Chen, "Modeling and characterization of gate oxide reliability," *IEEE Trans. Electronic Devices*, vol. 35, Dec. 1988, pp. 2268–2278.

[59] ITRS, "International technology roadmap for semiconductors: 2003 edition," tech. rep., International Technology Roadmap for Semiconductors, 2003. http://public.itrs.net.

VITA

Jung Yoon Hwang received the B.S. degree in physics from the Military Academy, Seoul, Korea, in 1993, and the M.S. degree in industrial engineering from Texas A&M University in 2001. Before he resumed his study in 1999, he served in the Republic of Korea Army for 5 years as a platoon leader and a staff in the armor brigades. His research interests include yield and reliability modeling and the quality analysis of integrated circuits and nano systems. His address is as follows: The University of Tennessee, 124 Perkins Hall, Knoxville, Tennessee 37996-2000.