Decision Support System to Predict the Manufacturing Yield of Printed Circuit Board Assembly Lines

Felipe Helo

A Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Industrial and Systems Engineering

Dr. John Kobza, Chair
Dr. Kimberly P. Ellis, Chair
Dr. Joel Nachlas

April 27, 2000
Blacksburg, Virginia

Keywords: Surface Mount, Yield, Fault Coverage, Defect-level,
Printed Circuit Board design

Decision Support System to Predict the Manufacturing
Yield of Printed Circuit Board Assembly Lines

Felipe Helo

ABSTRACT

This research focuses on developing a model to predict the yield of a printed circuit board manufactured on a given assembly line. Based on an extensive literature review as well as discussion with industrial partners, it was determined that there is no tool available for assisting engineers in determining reliable estimates of their production capabilities as they introduce new board designs onto their current production lines. Motivated by this need, a more in-depth study of manufacturing yield as well as the electronic assembly process was undertaken. The relevant literature research was divided into three main fields: process modeling, board design, and PCB testing. The model presented in this research combines elements from process modeling and board design into a single yield model.

An optimization model was formulated to determine the fault probabilities that minimize the difference between actual yield values and predicted yield values. This model determines fault probabilities (per component type) based on past production yields for the different board designs assembled. These probabilities are then used to estimate the yields of future board designs. Two different yield models were tested and their assumptions regarding the nature of the faults were validated. The model that assumes independence between faults provided better yield predictions.

A preliminary case study was performed to compare the performance of the presented model with that of previous models using data available from the literature. The proposed yield model predicts yield within 3% of the actual yield value, outperforming previous regression models that predicted yield within 10%, and artificial neural network models that predicted yield within 5%.

A second case study was performed using data gathered from actual production lines. The proposed yield model continued to provide very good yield predictions. The average difference with respect to the actual yields from this case study ranged between 1.25% and 2.27% for the lines studied. Through sensitivity analysis, it was determined that certain component types have a considerably higher effect on yield than others. Once the proposed yield model is implemented, design suggestions can be made to account for manufacturability issues during the design process.

ACKNOWLEDGEMENTS

Many people have contributed to the completion of this thesis. I am particularly thankful to those of you that have spent time guiding me and supporting me throughout these two years.

Dr. John E. Kobza and Dr. Kimberly P. Ellis have been my co-chair advisors while I have been working on this research. They invited me to be a part of their research team during my first semester as a graduate student. I have learned a lot during these years. Working with them has been a very positive experience in my career.

Dr. Joel Nachlas has also served as a member of my research committee. As an undergraduate student unsure of whether or not to continue in school, Dr. Nachalas encouraged me to stay in school and I appreciate his advice. I want to thank Dr. Nachlas for his guidance in the development of this model from beginning to end. His references were always accurate and extremely useful.

Finally, I would like to thank our corporate sponsor for their support throughout the project. Working with such a well known company that is committed to improving its operations has truly been an honor. I am particularly thankful to the engineers in the European facility that provided the data for me to use in the case study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1. Overview of Electronic Assembly Process

TVs, computers, stereos, phones and many other electronic products that characterize life at the end of the 2Oth century contain printed circuit boards (PCBs). During the last two decades technological advances have given rise to PCBs of increasing complexity [Chevalier, 1997]. Boards that require hundreds or thousands of assembly operations are not uncommon in the manufacture of today's electronic products. Consumers of these products are also demanding higher quality and more reliable items [Joseph, 1990]. Under these circumstances, greater control is necessary over the manufacturing processes used to produce PCBs.

In the past, Through-Hole assembly was the most common technology used to produce PCBs. Currently, Surface Mount Technology (SMT) is more widely used because it is faster and more precise [Moyer, 1997]. With Through-Hole assembly the components are inserted through actual holes in the board, while with SMT assembly, the components are placed on a solder adhesive at pre-specified locations on the board.

The manufacturing process for producing PCBs can be generalized into three main steps: paste application, pick-and-place, and reflow. For paste application, each board has a unique stencil; a "squeegee" pushes soldering paste through this stencil onto the bare board where components will be located. The key is to place the right amount of solder in the right place on the board. Too much paste, for example, could create a solder short that would cause the board to fail. After the paste is applied a pick-and-place

machine that was previously loaded with components, picks the needed component, orients it correctly, and places it at the specified location on the board. Finally, the board is preheated below solder reflow temperature. Once a uniform temperature is achieved across the board the temperature is increased again so that reflow (the process of melting the solder paste on the printed circuit board to provide electrical connectivity and hold components) occurs uniformly [Orsejo, 1998].

The large number of operations that are required for producing PCBs make the process very complex and therefore high failure rates are expected [Hroundas,1986]. Failure rates can be quantified by using the manufacturing yield of the process.

1.2 Manufacturing Yield and Faults

Manufacturing yield can be defined as the ratio of the number of good items produced to the total number of items produced. The difference between the total number of items produced and the number of good items produced accounts for the number of faulty items. It should be noted that more than one fault might be present in a faulty item.

Faults as defined by Hewlett Packard are “unacceptable deviations from the norm” [Orsejo, 1998]. Testing/inspection is the means through which faults are observed. The objectives of testing/inspection are to:

- detect the faults,
- diagnose the faults,
- provide process control information, and
- ensure confidence in the product.

In the printed circuit board industry there are three major fault classes: manufacturing faults, performance faults, and specification faults [Orsejo, 1998]. Missing components and insufficient solder are examples of some of the most common manufacturing faults. Performance faults occur when the individual components are within tolerance but the accumulated deviation of multiple components causes the board fail. Finally, specification faults account for the cases in which a board fails under special design conditions. Failing to meet environmental, contractual, or regulatory requirements are examples of specification faults.

Currently, the process through which fault information is used to predict manufacturing yield is not well defined. In particular, there are no specific tools that have been developed solely for this purpose.

1.3 Motivation of Research

This research focuses on developing a model to predict the manufacturing yield of printed circuit boards. It is motivated by the need for a tool to assist the manufacturers of electronic products in determining reliable estimates of production capabilities as they introduce new PCBs into their production lines.

CHAPTER II

PROBLEM DESCRIPTION

2.1 Problem Overview

This thesis develops a methodology and uses it as the basis for a model to predict the manufacturing yield of PCBs. Two different approaches are investigated and combined:

- process yield estimation, and,
- board design yield estimation.

Process yield estimation examines the process capabilities and estimates yield based on a predetermined fault spectrum. It incorporates concepts used in the computation of yield for integrated circuits (ICs). Board design yield estimation considers the effect that design specifications and characteristics have on manufacturing yield. Previous research has not combined process yield estimation and board design yield estimation.

Combining both methods into a single model provides more information regarding the causes of the faults and a more representative fault spectrum. Consequently, the resulting model will provide better estimates for the manufacturing yield.

2.2 Problem Statement

This section describes the manufacturing yield estimation problem. The objective is to develop a model that will predict the manufacturing yield for a given PCB that is assembled on a specific production line. The yield model is expected to be used to:

- monitor current PCB production,

- assist in the introduction of new products on current production lines,
- assist in the design of new products, and,
- compare the performance of similar production lines by comparing expected and actual yields.

The problem description is presented in Figure 2.1.

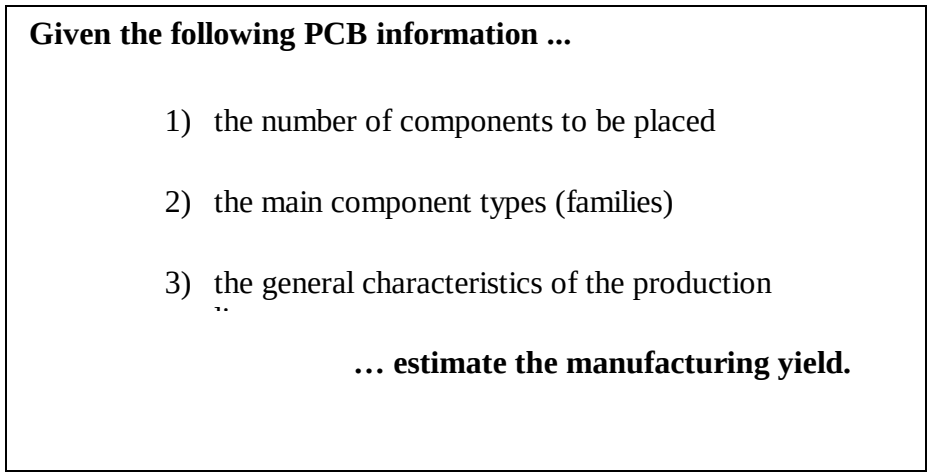

Figure 2.1 Manufacturing Yield for a PCB Production Line Problem Statement

2.3. Research Strategy for Addressing the Problem

The research approach for addressing the yield estimation problem involves the following activities:

- Review the related literature.
- Develop an approach for combining process and board design characteristics to estimate manufacturing yield.
- Conduct a preliminary case study to compare predicted yield results with actual results from current production lines.
- Develop and refine the implementation methodology for predicting the yield of new board designs as they are introduced into current production lines.
- Summarize findings and recommendations.

A brief overview of these activities is presented in this section.

The literature review develops the concept of manufacturing yield and groups the existing research into main categories. In addition, it provides a base for the development of several concepts used to generate and implement the yield model.

The yield model considers process capabilities and board design characteristics (number of components, component types, etc.), and numerically quantifies this information to determine component type fault probabilities. Once determined, these probabilities are used to predict the yield for new board designs.

Through various case studies using actual production data, the model is tested, validated, and refined. Conclusions are then drawn regarding the validity and applicability of the proposed yield model and its underlying assumptions. Finally, the implementation of the validated yield model is presented.

CHAPTER III

LITERATURE SURVEY

This chapter presents the literature related to the manufacturing yield of PCBs. The research done in this field can be divided into three main areas: *process modeling*, *board design* and *testing*. *Process Modeling* develops a model (usually statistically based) to mathematically predict the manufacturing yield of a given process. *Board Design* empirically predicts the manufacturing yield by looking at the design parameters and characteristics of the PCB. Finally, *PCB Testing* addresses the close relationship that exists between yield estimation and the percentage of faults that are detected when testing. The first three sections of this chapter summarize the most relevant literature within each of these three research areas. The last section presents a summary of the three as they relate to manufacturing yield.

3.1 Process Modeling

3.1.1. Integrated Circuits

In the late 1970's and early 1980's the development of more complex integrated circuits chips by the electronics industry made it clear that a method for quantifying the low yields that were being obtained needed to be developed. In particular, defect level and fault coverage became the key elements researchers examined to determine the yield of a particular process. Defect level is the percentage of chips delivered with undetected faults, and fault coverage is the percentage of faults that will be detected during board test [Aas, 1988]. In 1978, Wadsack [Wadsack, 1978] quantified the relationship between the two as linear. In 1981, Wadsack [Wadsack, 1981] redefined his model and introduced a

two-parameter gamma distribution to model the differences in defect density across a chip. Later that year, Williams and Brown [Williams, 1981] presented a paper in which a nonlinear relation was derived for defect level versus production yield and fault coverage. They assumed that a given chip had a given number of faults and that faults were independent of whether or not other faults had occurred. Based on these assumptions, they first derived an expression assuming uniformly distributed faults and using the binomial distribution (Figure 3.1.).

n = Total # of processes of type i

p = Probability of fault of type i occurring

DL = Defect level

Y = Yield

T = Fault Coverage

Probability Board No Faults $= (1-p)^n$

Probability of k Faults $= \binom{n}{k}(1-p)^{n-k}\,p^k$

Probability of accepting a board with k faults when testing m of the n:

$$\binom{n}{k}(1-p)^{n-k}\,p^k\,\frac{\binom{n-k}{m}}{\binom{n}{m}} = \binom{n-m}{k}(1-p)^{n-k}\,p^k$$

Probability of accepting a board with one or more faults when testing m of the n:

$$Pa = \sum_{k=1}^{n-m}\binom{n-m}{k}(1-p)^{n-k}\,p^k = (1-p)^m - (1-p)^n$$

$$DL = \frac{Pa}{(1-p)^n + Pa} = 1-(1-p)^{n-m}$$

$$Y = (1-p)^n$$

$$DL = 1-Y^{(1-m/n)} = 1-Y^{(1-T)}$$

Figure 3.1. The Williams and Brown Model

The Williams and Brown model became extremely popular among researchers in the field. Multiple papers cite their work as a key milestone in the development of a mathematical model for the estimation of yield. In 1982, Agrawal [Agrawal, 1982], presented a variation of the Williams and Brown model incorporating the average number of faults on a chip and the yield of good chips. This model calculates the rejection rate as a function of the fault coverage.

Williams, working as a senior technical staff member for IBM corporation during the 1980's and early 90's, developed various modifications of the Williams and Brown model. In 1985, he presented a paper [Williams, 1985] in which the relationship between defect level and random patterns was used to determine accurate test procedures in a self-testing environment. Park [Park, 1989] presents, a way for statistically measuring delay-fault coverage. Park then generalized this model by considering a testing strategy which is determined by the defect level for a given set of faults [Park, 1992]. Corsi [Corsi, 1993] uses conditional probability to simplify the assumption made in the original Williams and Brown model regarding equiprobable faults. This model incorporates a generalized weighted fault coverage parameter that accounts for the non-equiprobable faults. Finally, Sousa [Sousa, 1996] extends the model once more to incorporate the concepts of non-equiprobable faults into defect-level calculations. This model computes the probability of occurrence of a given fault, provided that a critical area for the fault can be defined. The critical area can be determined by analyzing "typical" IC process line defect statistics. The authors concluded that the defect level depended on the critical areas associated with undetected faults and fault densities.

In the 1980's, C. H. Stapper, also working for IBM, did extensive research that involved applied statistics and simulation to model manufacturing yield of integrated circuits. Stapper [Stapper, 1989] presents two simulation programs that were developed at the Esssex Junction, Vermont IBM facility. The first program generates a negative binomial distribution to represent the frequency distribution of the number of faults per chip. The second program simulates clustered fault locations on a map using a radial Gaussian probability distribution.

Following Stapper's research, Aas and Minh [Aas, 1989] from the Norwegian Institute of Technology developed various simulation experiments to study multiple fault coverage versus single fault coverage. They defined multiple fault coverage as the event where the combined effect of *k* faults leads to fault detection although each individual fault was not detected. They also defined fault masking as being the event in which one or more of the *k* faults will lead to single fault detection, but the combined effect of the *k* faults will lead to simultaneous fault masking. They used the Poisson and Poyla distributions as fault distributions to create their model and incorporated conditional probabilities to determine the probabilities for multiple fault coverage. In their study, they found that fault exposure and fault masking are high for single fault coverage and that it is less pronounced for multiple fault coverage.

Due to the increased research in the field of fault coverage for integrated circuits the obvious question is: how much fault coverage is enough? In 1989, B. Henshaw [Henshaw, 1989], working for NCR Corporation, used a cost model to answer this question. In his research, Henshaw developed an equation to calculate the defects in parts per million, and fed this information into a cost model. He concluded that for

certain typical conditions the economic level of fault coverage should be as close to 100% as possible.

3.1.2. Printed Circuit Boards

The research presented so far has been related to integrated circuits. As printed circuit boards increased in complexity, manufacturing yield of PCBs became an issue and a topic of study for researchers [Chevalier, 1997]. Most of this research has been done at major corporations that produce electronic products, such as Hewlett-Packard, Ericsson, Motorola, ATT, and IBM. Electronic products found in today's market typically contain one or more PCBs; and, therefore, the production of PCBs will directly affect the production of electronic products. Due to the extreme competition that exists among some of these companies, much of this research has not been published. The following section presents some of the most important published research that relates to estimating the yield of PCBs.

Joseph, Watt and Wigglesworth [Joseph, 1990] describe a project focused on the assembly and test of electronic sub-assemblies at Digital Equipment Corporation. Their objectives were to:

(1) evaluate the impact of different strategies for assembly, test, and repair processes;

(2) evaluate the relationship between incoming quality levels, assembly quality levels, test screen strengths, cycle time and capital requirements; and,

(3) understand the relationship between test development, assembly process development and volume process costs.

The resulting tool addressed yield analysis and detailed capacity planning using fish-bone analysis, IDEF modeling, and simulation. The design process is shown in Figure 3.2. The Fish-Bone Analysis was used to analyze defects. Charts were created for 20 distinct defect types. Figure 3.3 shows an example for the poor bonding fault. These charts were used to identify the defect information of the fault model including:

- incoming defect rates,
- process-induced defect rates,
- sources of manufacturing defect,
- defect capture points,
- defect capture rates.

The IDEF functional modeling technique was used to model the manufacturing process and to provide a conceptual model for use in developing the simulation model. The simulation model was implemented in SLAM II and provided system performance estimates of:

- work in process (WIP),
- utilization,
- equipment requirements (number of machines, staffing, buffers),
- cycle time,
- throughput,
- defect escapes, and
- product yield.

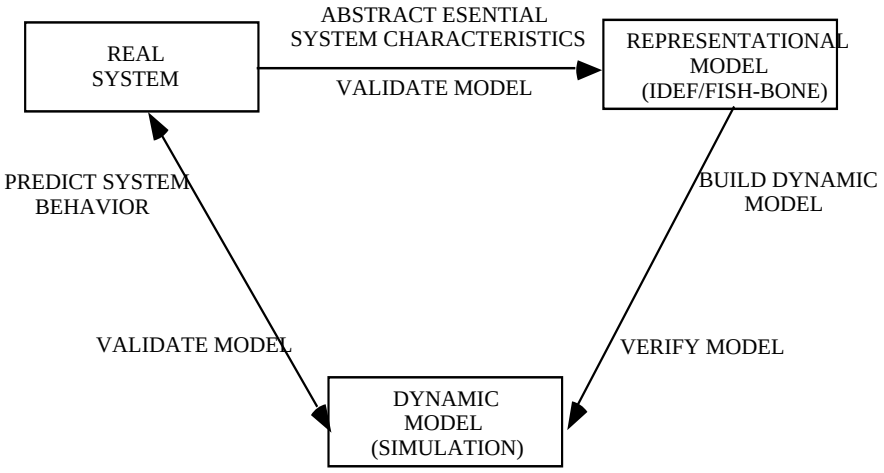

Figure 3.2. Life-Cycle of Analysis

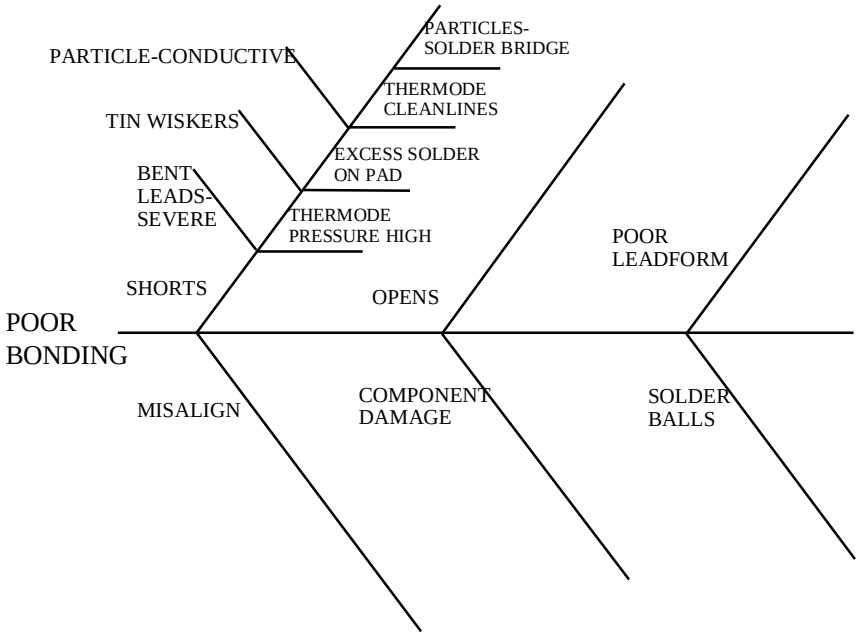

Figure 3.3. Fish-Bone Chart for Fault POOR BONDING

Working for a private contract manufacturer, Clark, White, Leany and Wycherley [Clark, 1995] developed a quality modeling system (QMS) for predicting the yield of assembly and test processes in the printed circuit board assembly industry. The system

looks first at the cause of the fault; then, breaks it down into two main fault categories: assembly and component. Assembly faults are related to the package of a component, whereas component faults are related to the technology of the component. They used actual production line performance data and collected the defect rates using the fault collection system MARQUIS. Once information regarding the specifics of the particular board was entered, MARQUIS collected the fault information, grouping it by board type and fault location. Information regarding the type of component at each location was held on another system called SPEAR which is used for assembly planning. Both packages share INGRES databases making information easy to merge.

In order to correctly track the source of the fault QMS looks at three particular data requirements:

- part number,
- processor (where the board is being assembled), and
- fault type.

Once faults are identified, QMS divides them into either assembly faults or component faults. Assembly faults are divided into those faults that cause:

- short circuits,
- open circuits,
- non-functional faults, and
- visually unacceptable faults.

Defect rates were generated for each one of these assembly faults. The authors assumed component faults were related to the type of electronic functions that they perform. This assumption was made after analyzing historical data and concluding that

the defect rate was partially dependent on the components' complexity and technology. They also categorized the boards into three different complexity levels: simple, complex and very complex. The distinction between the three was based mainly on the component density of the board which was computed by taking the ratio of the area required to mount all the components over the total area available for mounting components on the board.

QMS was developed using EXCEL and ACCESS. The ACCESS software stores the information regarding boards and faults and serves as the data entry interface. The EXCEL spreadsheet was used to extract the relevant information from the ACCESS database and to perform the needed calculations. Computing the yield was done using the Poisson distribution (a simplification of the binomial distribution from the Williams and Brown model). The model was tested using forty different boards already in production and the results showed that the model was accurate to within 5% of the actual assembly yield. The QMS model has been used since 1994 and, according to the authors, provides useful information for the manufacturers to set production targets for volume and to monitor production.

In 1995, working for Tandem computers, Collins, Tuttle and Tuttle [Tuttle, 1995] presented a paper in which the Williams and Brown model as well as the Agrawal model were used for predicting the yield of PCBs. They describe a method for quantitatively linking defect levels, test effectiveness, and failures to both the customer and the producer. The goal of the methodology presented was to quantify the financial impacts of specific board requirements. It looked at past mistakes and aimed at coupling manufacturing capabilities with future board designs.

Given that manufacturing yield directly depends on the number of faults in the process, Linn [Linn, 1996] studied the placement errors for Surface Mount Technology. He determined that the placement process errors depended on three characteristics:

- the components used,
- the design characteristics of the PCB, and
- the placement machine.

Errors at the component level are mostly related to geometric specifications of the leads. Errors at the PCB level usually are caused by dimensional variations and location offsets. Finally, placement equipment errors are caused by either positioning accuracy, rotational accuracy, or improper vision alignment. Linn derived an equation for computing the probability of successfully placing a component, taking into account the characterization of errors previously described.

3.2 Board Design

The number of fault-free PCBs depends not only on the quality of the components used and the accuracy/capabilities of the processes, but also on the design of the board. Li, Mahajan and Tong [Li, 1993] found that as the design complexity of the board increases, the yield decreases. Faults are more likely to take place in more complex designs because the probability of error increases as:

- the number of components increases, and,
- the complexity of the assembly process increases (placement precision, tolerances, etc.)

Based on this idea, research has been done in which the design of the board becomes the key element in predicting the manufacturing yield of PCBs.

The most important study that uses this concept to predict the yield was done by Li, Mahajan and Tong [Li, 1994] from the University of Colorado and the AT&T Engineering Research Center in Princeton, NJ. In this study, 30 different printed circuit boards running on two different production lines were studied over a period of one year. The information collected included data related not only to actual process yields, but also data related to board design and component characteristics.

Li, Mahajan and Tong categorized the design factors that affected assembly yield into five major groups:

(1) number and package types of the components on the board,

(2) number of different types of leads,

(3) density,

(4) component layout, and

(5) board geometry.

The authors used a regression model and an artificial neural network model to establish the relationship between the design factors and the assembly yield. The regression model was first developed taking into account all of the different characteristics of the particular components placed on each board. This model was then simplified using ANOVA in order to determine the critical components that had the greatest effect on yield. Data generated by the regression model was within 10% of the actual production data. The artificial neural network model was developed by dividing the collected data into training data and testing data. The model requires data in order to

learn. Eventually, no significant improvement is gained using additional data, and the model reaches an optimum network architecture that is tested using the test data. This model produced results within 5% of the measured values for the manufacturing yield.

Holden [Holden, 1995], working for Hewlett-Packard, developed a different model for computing the first pass yield by introducing the concept of *complexity index*. Holden's motivation was to predict the relative manufacturing costs of bare PCBs; however, his methodology is general enough that it could be applied to populated PCBs as well. The model looks at the bare board and computes a complexity index based on the board's design characteristics. This index is a ratio of parameters and characteristics of the PCB. The first pass yield is then computed by taking the log of the normalized complexity index elevated to a given power, and taking the negative exponential of it (Figure 3.4).

$$\text{Complexity Index } (CI) = \frac{(\text{area})(\text{no. holes})(\text{T})(\text{no. layers})(\text{T}_0)}{(\text{A})(\text{H})(\text{min. trace width})(\text{L})(\text{min. tol.})}$$

where $\text{A}, \text{H}, \text{L}, \text{T}$ and T_0 are characteristics of a PCB with 100% Yield

$$\text{First Pass Yield \%} = \frac{100}{\exp\{(\log \frac{CI}{A})^{B}\}}$$

Figure 3.4. Holden's First Pass Yield for Bare Boards

The normalization constant, as well as the exponent used, are determined by regression analysis. Holden tested his model on the assembly of bare printed circuit boards and concluded that as the complexity index of the boards increased, the first pass yield decreased. The rate at which the yield decreased depended on the values of the

constants determined by regression analysis which in turn depended on the design of the board.

3.3 PCB Testing

PCB testing is a broad area of research. The following section presents only the PCB testing literature realted to manufacturing yield. It should be noted that testing is an open area for future research that could be explored in much more detail.

Testing a PCB has both benefits and costs associated with it. The benefits of testing include improved quality control and process feedback. Some of the costs include test time, which affects test station throughput, hardware complexity, which affects equipment costs, and programming time. The quality of the test process is directly related to the percentage of faults that will be detected by the test (i.e., fault coverage). Hotchkiss, working for Terdayne, points out that the key tradeoff in test planning is between fault coverage and programming time. In essence, there is a point at which the additional increase in fault coverage does not justify the increase in programming time. The research showed that the decision depends on board volume and the number of board types to be tested. For lower volumes, the additional investment in programming time might not be cost effective [Hotchkiss, 1980].

Hroundas [Hroundas, 1990] explained that, to remain competitive in the 1990's, the manufacturers of PCB's needed to look, in particular, at the quality of electrical testing. He states that in order to achieve parts per million (ppm) failure levels, the manufacturer must increase not only the control over the processes, but also the quality of the testing equipment. According to Hroundas, fault spectrums are constantly changing and, therefore, testing equipment needs to adjust guarantee high quality products. In

summary, the high quality of the PCB's is directly dependent on the quality of the electrical test it receives; and the quality of the electrical test depends on the testing equipment's fault coverage.

Complementing Hroundas' research, Millman [Millman, 1994] working for Motorola, explained how increasing test coverage could potentially have the same effect on quality as improving the manufacturing yield (assuming that a reasonable yield estimate is available). Millman compared the two from an economic perspective and concluded that increasing the test coverage has a greater impact on quality for a lower cost than similar increases in yield. He compared increasing test coverage and increasing yield to determine the effect on the quality level. The quality level is the ratio of the number of defect free parts that pass the test over the total number of parts that pass the test. Millman used the Williams and Brown model as well as the Agrawal model to compute the yield. For the fault coverage, he points out that the key element is to accurately match the fault model used with the faults actually occurring. In particular, to properly increase the fault coverage, the type of failure must be identified and the conditions under which it occurs must be determined. In his model, Millman described that, if the test coverage is greater than a certain value (computed as a function of the yield and the cost of testing equipment), then a greater increase in quality level will result from an increase in test coverage than from an equal increase in yield.

In 1992, working for Hewlett-Packard, Tegethoff, Figal, and Hird [Tegethoff, 1992] presented a Design for Test (DFT) model for PCBs. Their model combines test effectiveness and the board fault spectrum to determine the yield and the cost for each test step. The model inputs are fault spectrum, test effectiveness, specifics of the repair

process, and product test costs. The main outputs of the model are product yield and product test costs. The model computes the yield by dividing the faults into component faults and assembly faults, and then uses the Poisson approximation of the binomial distribution to compute the overall yield. The DFT model assumes:

(1) assembly and component defect rates are mutually independent,

(2) test effectiveness is a measure of finding defects,

(3) component category defect rates are mutually independent,

(4) in-circuit test finds multiple defects per test category, and

(5) functional test finds defects per cycle.

The authors point out that by doing sensitivity analysis on any or all of the inputs, the model will assist in developing the test strategy. Implementation of the model has given cost and yield predictions within 20% of the actual manufacturing process. The authors consider these yield predictions to be acceptable due to the simplicity of implementation and usage of the model.

Also working for Hewlett-Packard, Orsejo [Orsejo, 1998] presented several test strategies for the manufacturing of PCBs by classifying the major fault classes in Surface Mount Technology. His paper presented a fault spectrum from data collected for over 40,000 different boards and 74 million solder joints. The fault spectrum (see Figure 3.5) accounted for eight different fault types. He then looked at several testing strategies and matched the fault spectrum to the fault coverage claimed by each individual test strategy.

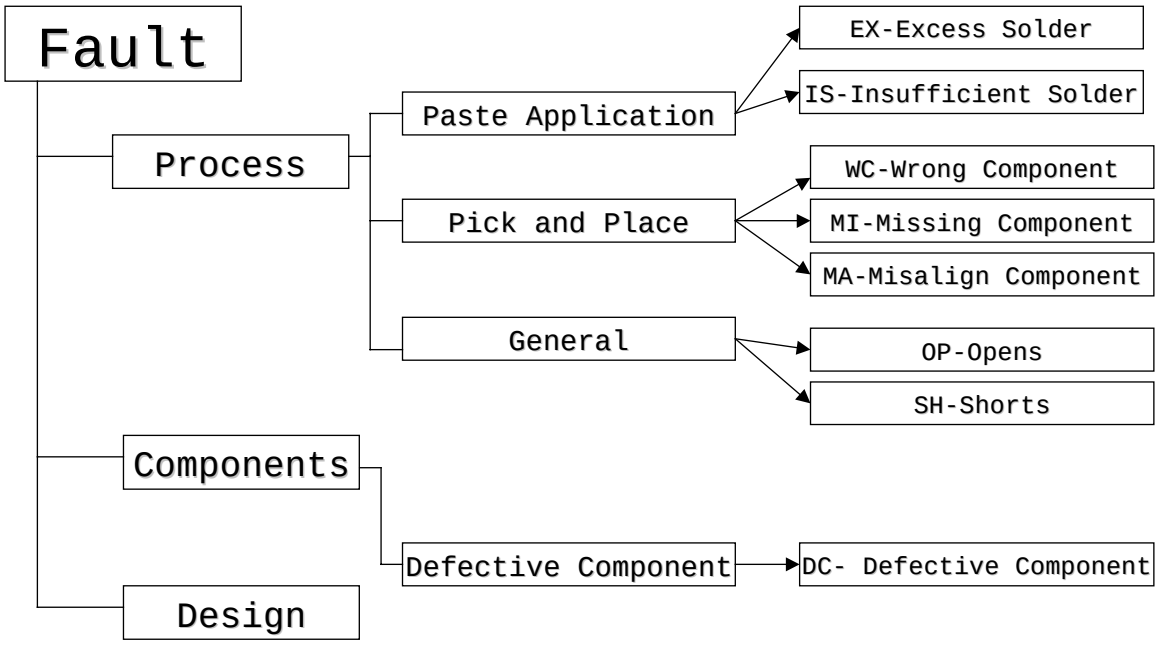

Figure 3.5. Orsejo's Fault Spectrum

Orsejo analyzed the following testing strategies:

- manual inspection,
- laser inspection,
- automated optical inspection,
- two-dimensional automatic X-ray inspection,
- three-dimensional automatic X-ray inspection,
- flying probe, manufacturing defect analyzer,
- in-circuit test and boundary-scan.

As expected, the study showed that the best economic strategy for the production of PCB is to have the highest possible fault coverage as early as possible. He concluded that no one testing strategy would fulfill all of the fault spectrum requirements. An appropriate

combination of test strategies, therefore, is necessary for high yield, high quality, cost-effective products.

3.4 Summary of Literature Review

Due to the continuous increase in demand for higher quality electronic products, estimating yield for printed circuit boards has been a topic of research since the late 1970's. The manufacturing yield models have mostly been based on models originally developed for integrated circuit production lines. The Williams and Brown model (using the Poisson approximation of the binomial distribution) is the most popular model that characterizes the nonlinear relationship between defect level and yield. Studies have also been done in which the manufacturing yield has been computed by looking only at the design characteristics of the board. Finally, the characteristics of the testing equipment as well as the impact of appropriate testing strategies on manufacturing yield have been analyzed.

In summary, the literature review presented two different modeling methodologies for estimating manufacturing yield: process modeling and board design. These methodologies were developed independently and approach yield estimation in entirely different ways. A main gap in the literature is that the two methodologies have not yet been combined into a single model. The following chapters present the development of such a model showing its implications, applicability, and usefulness.

CHAPTER IV

YIELD MODEL DEVELOPMENT

The goal of the manufacturing yield analysis presented in this research is to combine *process modeling* and *board design* into a single yield model. This can be accomplished by: (1) looking at research that has been done in both of these fields, (2) determining similarities between the two, and (3) formulating a procedure that combines them into a single model. Three areas need to be explored in the development of this single model:

- model assumptions,
- data requirements, and
- model formulation.

The following sections explain each one of these areas in greater detail.

4.1. Model Assumptions

The primary assumptions incorporated in the yield model are briefly described in this section.

1. Process Modeling Assumption. As discussed in Chapter III, the Williams and Brown model has been the most widely used model to predict yield. This model has not only been used for IC applications but it has also been used for PCBs (see Section 3.1.2). The main assumption for the Williams and Brown model is that the probability that a fault occurs is independent of the number of other faults on the board. For the particular case in which the model aims to predict the yield of PCBs, this assumption can be rewritten as:

The probability of a fault occurring on a PCB is independent of the faults already on the board.

2. Board Design Assumption. As was presented in Section 3.2, Li, Mahajan, and Tong [Li, 1994] showed that two of the main design factors that affect assembly yield are the total number of components and the various types of components on the board. A component/lead must fall within one and only one of the predetermined component/lead types. This assumption avoids cases in which a component is counted more than once by the model because it falls within more than one component type category. It can be stated as:

The total number of components in a board must equal the sum of all of the individual number of components/leads for each component/lead type.

3. *Combination Assumption*. In order to incorporate the design elements (total number of components/leads to be placed, and the component/lead types), into the Williams and Brown model an additional assumption must be made. In particular,

The fault rates for the component types are mutually independent.

4.2. Data Requirements

The two modeling methodologies require certain data in order to estimate yield. *Process modeling* uses a given fault probability (or fault spectrum) and computes the expected yield based on it. Fault probabilities are the main data input for this model. On the other hand, *board design* uses available data regarding board characteristics and past

yields to determine regression coefficients to be used for predicting yields. The data inputs for this model are the number of components/leads, component/lead types, board design characteristics (geometry, density), and yield history for old board designs. A procedure for obtaining the fault spectrum (input for *process modeling*) from the input data for *board design* is described in the following model formulation.

4.3. Model Formulation

Fault probabilities required to estimate yield for *process modeling* are determined through observation. The main drawback of this approach is that this data is difficult to collect because it is not always available. For example, information regarding the quality of supplier products as well as the effectiveness of in-house processes must be known. The yield model formulated in this section overcomes this drawback by providing a methodology for estimating fault probabilities using board design information and historical yield information.

The first step in the development of this model is to determine why fault probabilities are required for *process modeling* yield estimation. Recall that for *process modeling* the Williams and Brown model (shown in Figure 3.1) is the most commonly used model. This model uses the binomial distribution to account for the probability of fault occurrence. A slight modification of this model follows.

Let,

n= Total number of processes

p= Probability of a fault occurring

k= Total number of faults on the board

and, let the fault rate λ be,

$$\lambda = n \cdot p \ .$$

The binomial distribution for the number of faults can be expressed in terms of λ as,

$$b[k; n, p] = \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^{k}\left(1-\frac{\lambda}{n}\right)^{n-k}, \qquad k= 0,1,\ldots,n.$$

It should be noted that the Poisson approximation to the binomial distribution provides sufficient accuracy because *n* is large and *p* is small for most cases that will be encountered [Tegethoff, 1992]. For practical reasons this Poisson approximation will be used throughout the development of the yield model. Therefore, the fault probability of having *k* faults on a printed circuit board can be expressed as:

$$P(k \text{ faults}) = \frac{e^{-\lambda}\lambda^{k}}{k!} \ . \tag{4.1}$$

According to the definition of manufacturing yield (Section 1.2) an item is either defective or non-defective depending on whether or not faults occur. Yield is the probability that no faults occur resulting in good items. Therefore the Poisson model (4.1) predicts yield as:

$$Y = P(k = 0) = \frac{e^{-\lambda}\lambda^{0}}{0!} = e^{-\lambda}. \tag{4.2}$$

Fault probabilities need to be known in order to determine λ (since λ =np), which in turn is necessary to estimate yield using Equation 4.2. Note that Equation 4.2 represents the yield of a given process and does not yet account for specific design characteristics of the board. To incorporate these elements the model can be reformulated as follows:

Let

p_i = fault probability for components of type *i*.

n_i = number of components of type *i*.

λ_i = fault rate for component type *i*.

N = total number of component types

where

$$\lambda_i = n_i \cdot p_i \quad .$$

Using the *board design assumption* (Assumption #2), the total number of components for a given board will be represented by the sum of the number of components per component-type category, i.e.,

$$\text{Total no. of components per board} = \sum_{i=1}^{N} n_i \ .$$

The *combination assumption* (Assumption #3), states that the processes are mutually independent of each other. Under these assumptions the yield model can be further generalized to incorporate *board design* specifications as follows:

$$Y = \prod_{i=1}^{N} e^{-\lambda_i} = e^{-\sum_{i=1}^{N} \lambda_i} = e^{-\sum_{i=1}^{N} n_i * p_i} \quad . \qquad (4.3)$$

As shown, this model requires various design inputs such as the total number of component types (N) and the number of component for each type (n_i). It also requires a component-type fault spectrum (p_i *values*) in order to predict yield. The remainder of this chapter will present a methodology for computing this fault spectrum using the yield history for previous board designs.

As mentioned in Section 4.1.2, an important data requirement when using a *board design* approach is to have actual yield data from previous board designs. When this is the case, the unknown values of λ_i can be estimated by formulating an optimization model to fit these data. In particular, the objective is to determine the component-type fault probabilities by looking at the past performance of production lines.

The objective of the optimization model is to minimize the difference between the predicted and actual yields for different board designs, by optimizing the component-type fault spectrum (p_i *values*).

Let

NB = *total number of different boards,*

j = *board number,* $j=1,\ldots,NB$,

y_j = *actual yield for board j,*

N = *total number of component types,*

$n_{j,i}$ = *number of components of type i for board j.*

Using Equation 4.3, the predicted yield for board *j* can be expressed as

$$\hat{Y}_j = e^{\sum_{i=1}^{N} \lambda_i} = e^{\sum_{i=1}^{N} p_i n_{j,i}}.$$

The objective is to minimize the difference between the actual and the predicted yields for a given board design. This can be expressed as:

$$\left| \frac{y_j - e^{\sum_{i=1}^{N} p_i n_{j,i}}}{y_j} \right|.$$

The overall optimization model for *NB* different board designs can be expressed as shown in Figure 4.1.

Given

NB = total number of different boards
j = board number, j = 1, ..., NB
y_j = actual yield for board j
N = total number of component types
$n_{j,i}$ = number of components of type i for board j

Find

p_i = fault probability for components of type i

Minimize

$$\sum_{j=1}^{NB} \left| \frac{y_j - e^{-\sum_{i=1}^{N} p_i n_{j,i}}}{y_j} \right|$$

Subject to

$0 \le p_i \le 1$.

Figure 4.1. Optimization Problem for Poisson Yield Model

After the optimization model is solved, the component-type fault spectrum can be used to estimate the yield of future PCB designs. This can be done by following a simple two step process: 1) group the components of the future board design into the current component-type categories for which the p_i values have been estimated, and 2) predict the yield by solving Equation 4.3.

In summary, this chapter has presented the basic formulation of a model to predict yield using concepts from *process modeling* and *board design* has been developed. In addition, methods for obtaining the data requirements for process modeling based on available board design information have been discussed. The following chapter looks at a preliminary case study in which the model is implemented to show its performance, as well as to validate *the process modeling assumption* made (Assumption #1).

CHAPTER V

MODEL IMPLEMENTATION AND VALIDATION

In the previous chapter, the mathematical foundation for the yield model was developed. This chapter looks at a specific case study in which the component-type fault spectrum is unknown and the Poisson optimization model described in Figure 4.2 is used to predict it. The effectiveness of the model is measured by the average percent difference between predicted and actual yield for various board designs. The same case study is performed under the assumption that faults are dependant, using the negative binomial model. Finally, a comparison of the results for both models gives insights regarding the validity of the fault formation assumptions.

It is also of particular interest to observe the effect that *board design* parameters have on the accuracy of the yield model. For this reason three different case scenarios are considered, each accounting for a different number of design parameters (i.e., total number of component-types considered).

5.1. Model Implementation

The first step for implementing the Poisson yield model is to determine the number of design parameters that need to be considered. The simplest case is the one in which every component is assumed to be the same and the same process is assumed to place all components. In this case, the yield model accounts for only a single process (a single component-type). Yield is then computed based on the probability of fault occurrence per component type. This is called a *single-variable* model since there is only one process parameter required to predict the yield.

A more interesting case is that in which the components/leads are divided into more than one category. For example, they can be divided according to the nature of the assembly process: through-hole or surface mount (see Chapter 1.1). This is a *two-variable* model since there are two parameters required for the model to predict the yield.

Finally, the most challenging case is that in which the number of processes represents the total number of different components types. This case is based on the specific lead and component classifications outlined in Appendix A. The addition of processes increases the number of parameters required to predict the yield making the model considerably more challenging to implement than the single-variable or the two-variable models. This model accounts for specific design characteristics of the PCB by looking at the probability of fault per individual component family.

The data used to create this preliminary case study was taken from Li [Li, 1993] and is presented in Tables 5.1 and 5.2. This case study consists of 30 board types each having similar types of components. Specific lead and component classifications are outlined in Appendix A. Figure 5.1 shows a breakdown of the case study into the three different scenarios at hand. Note that component-types are assumed to be independent processes, and *N* represents the total number of component types considered.

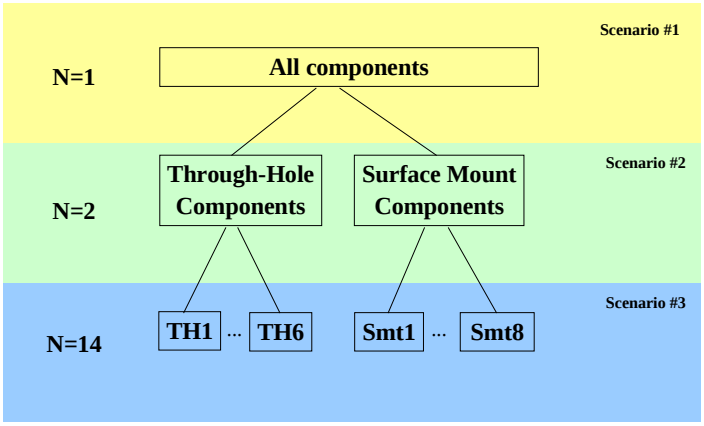

Figure 5.1. Breakdown of Case Study for Different Board Design Scenarios

Table 5.1 Descriptions of Design Factors

n	total number of leads
J	number of J leads
G	number of gull wing leads
Ct	number of terminations of chip components on the top side of the board
Cb	number of terminations of chip components on the bottom side of the board
SOTt	number of leads of SOTs on the top side of the board
SOTb	number of leads of SOTs on the bottom side of the board
DIP	number of leads of DIPs and SIPs
A	number of axial and radial leads
Nsth	number of non-standard through-hole leads
nshthd	number of non-standard hand inserted through-hole leads
nsmthd	number of non-standard machine inserted through-hole leads
DIPD	number of DIP and SIP components
JD	number of PLCCs
GD	number of gull wing SOICs
TOPr	density on top side of the board
BOTr	density on bottom side of the board

Table 5.2 Data for production lines of 30 different PCBs recorded at ATT presented in Li [Li, 1993]

Brd	Number of Leads																	Density		Actual
#	TH Leads						TH	SMT Leads				SMT	F.Pitch				N			Yield
	A	DIP	DIPD	Nsth	nshthd	nsmthd		J	G	Ct	Cb		SOTt	SOTb	GD	JD		BOTr	TOPr	
1	123	580	34	202	21	2	905	136	1024	50	1228	2438	96	78	59	2	3517	19.83	11.66	86.32%
2	323	240	18	288	8	32	851	204	314	60	778	1356	0	96	21	3	2303	16.42	5.25	80.64%
3	20	292	18	240	24	0	552	248	560	746	0	1554	108	0	32	14	2214	4.93	14.84	84.10%
4	95	28	1	136	13	8	259	208	512	442	0	1162	30	0	32	4	1451	2.31	10.63	88.23%
5	67	28	1	104	4	12	199	208	356	390	0	954	24	0	22	4	1177	1.78	8.71	88.44%
6	79	0	0	352	36	0	431	168	846	0	410	1424	0	99	44	3	1954	10.2	11	93.09%
7	73	64	1	246	11	0	383	237	1098	4	472	1811	9	0	61	7	2203	9.44	14.87	96.05%
8	134	610	30	356	13	21	1100	68	46	354	0	468	87	0	3	1	1655	11.52	6.03	90.30%
9	196	80	5	168	21	0	444	68	0	0	0	68	0	0	0	1	512	8.2	1.35	98.91%
10	696	672	35	346	24	2	1714	68	0	0	0	68	0	0	0	1	1782	17.91	0.74	96.65%
11	0	0	0	682	7	0	682	0	0	0	0	0	0	0	0	0	682	8.2	0	99.91%
12	278	160	6	110	10	8	548	0	0	0	0	0	0	0	0	0	548	10.67	0	97.61%
13	460	714	37	360	31	1	1534	68	0	0	0	68	0	0	0	1	1602	16.67	0.74	98.76%
14	436	646	34	331	11	0	1413	68	0	0	0	68	0	0	0	1	1481	16.15	0.78	95.70%
15	142	270	13	120	3	0	532	0	0	0	0	0	0	0	0	0	532	9.5	0	99.37%
16	336	242	13	184	11	0	762	0	0	0	0	0	0	0	0	0	762	13.61	0	98.14%
17	118	626	25	80	2	0	824	0	0	0	0	0	0	0	0	0	824	25.55	0	97.20%
18	184	610	30	356	13	21	1150	68	46	354	0	468	87	0	3	1	1705	12.15	6.03	91.29%
19	0	0	0	0	0	0	0	648	268	468	0	1384	7	0	34	15	1391	0	14.88	92.82%
20	0	0	0	0	0	0	0	2000	1678	20	572	4270	32	0	98	30	4302	5.11	33.3	79.49%
21	0	0	0	0	0	0	0	1724	1904	40	544	4212	19	81	113	50	4312	5.58	32.99	61.74%
22	469	464	19	68	1	0	1001	136	1888	0	0	2024	15	0	108	2	3040	8.3	18.21	82.57%
23	12	56	2	6	1	0	74	1564	2578	12	766	4920	0	0	148	43	4994	8.45	42.61	61.67%
24	2	0	0	40	1	0	42	80	184	6	134	404	0	3	10	3	449	16.62	25.2	88.48%
25	434	154	11	228	13	8	816	0	0	0	0	0	0	0	0	0	816	13.6	0	97.61%
26	74	206	13	168	11	0	448	0	184	422	0	606	24	0	10	0	1078	8	11.25	90.61%
27	65	0	0	472	39	0	537	460	668	8	516	1652	102	0	39	12	2291	13.73	13.46	81.43%
28	132	612	32	180	11	2	924	136	172	508	0	816	33	0	10	2	1773	10.15	9.23	88.84%
29	446	554	31	64	8	0	1064	272	0	0	0	272	0	0	0	4	1336	9.64	2.59	97.86%
30	60	280	13	78	3	3	418	504	1638	304	652	3098	20	0	96	12	3536	11.54	27	68.14%

5.2. Results for the Poisson Model Preliminary Case Study

The approaches used to solve the optimization problems for the three board design scenarios are presented in this section. The results for each scenario as well as relevant preliminary conclusions are also presented. A complete table of results for each individual scenario is included in Appendix B. The results in the appendix also provide the difference between the actual and predicted yields for each of the thirty board designs considered on an individual basis.

For the single variable model, the optimal solution can be determined since it is a two-dimensional solution space. A simple local search can be used to determine the minimum of this function by looking at the derivative of the objective function.

The solution to the single variable optimization problem (Scenario #1) for the data presented in Table 5.2, showed that on average this model predicts the yield of future board designs within 5.73% of the actual yield. The standard deviation of this solution is 5.51%. The maximum difference between the predicted yield and the actual yield is 23.44% with respect to the actual yield. Finally, the 95% confidence level for this model is 2.09%. The results from the single-variable model indicate that:

- the model is very easy to implement (no component-type assignments),
- the results do not provide reliable estimates since the mean and standard deviation of the difference between actual and predicted yields are high and,
- the model involves a simple computational process to reach an optimal solution (solution time negligible).

For the two variable scenario, where components are divided according to the nature of the manufacturing process (through-tole or surface mount), the solution can also be determined to optimality in a fairly simple way since it is a three-dimensional solution space. The approach used to solve this problem is similar to that used for the single case scenario, where the gradient of the objective function is used to solve the problem. At each iteration the search direction was taken as the negative gradient of the objective function (steepest descendent method).

The results of the two-variable model Scenario #2 showed that this model's predictions are on average within 3.76% of the actual yield with a standard deviation of 4.45%. In addition, the maximum difference with respect to the actual yield was found to be 17.43%, and the 95% confidence level to be 1.66%. The results from the two-variable model indicate that:

- the model is simple to implement although it must divide components into two main types,
- the results are considerably better than those obtained for the single-variable model suggesting that including more design parameters improves the performance of the model and,
- the model involves a fairly simple computational process to reach an optimal solution (solution time minimal).

Finally, for the multi-variable case (Scenario #3) the optimization model considered a total of 14 different component-types. The difficulty of the problem is directly related to the number of component-types involved. Each additional component-type adds a dimension to the solution space for the optimization problem making it considerably more difficult to solve. It is therefore expected that for situations

in which many component-types need to be considered, a heuristic may be necessary in order to simplify the solution approach. The compromise of using such an approach is that the solution obtained is not necessarily optimal.

In order to solve the multi-variable case (Scenario #3), a genetic algorithm heuristic approach was chosen. Genetic algorithms work by generating a "population" of possible solutions and ranking them in order from best to worst. The solutions are then paired and variables are swapped according to specified crossover and mutation rates to produce new solutions (offspring). If an offspring is good, it is inserted into the population, letting it evolve towards increasingly good solutions.

The results for the multi-variable case showed that the estimates of this model were on average within 2.91% of the actual yield with a standard deviation of 3.75%. The maximum difference with respect to the actual yield was reduced to 13.78%, and the 95% confidence level was reduced to 1.40%. The results from the multi-variable model indicate that:

- the model is considerably more challenging to implement than the other two scenarios (components must be grouped into fourteen different categories),
- the results more challenging to obtain (using genetic algorithms the problem took 17 hours to solve and the solution is not guaranteed to be optima); and,
- the model is significantly better than the two previous scenarios not only because the mean and standard deviation of the difference between predicted and actual yields are lower, but also because of the significant reduction in the range of the differences.

Table 5.3 summarizes the results of the three scenarios. In general, two important conclusions can be drawn from the results of this preliminary case study:

1. *The model becomes considerably more difficult to implement as more design parameters are considered.*
2. *Yield estimates are more accurate when more design parameters are considered.*

Table 5.3. Summary of Results for Poisson Model Preliminary Case Study

	Scenario #1	*Scenario #2*	*Scenario #3*
N	1	2	14
Average Difference for 30 PCBs	5.73%	3.76%	2.91%
Standard Deviation	5.51%	4.45%	3.75%
Maximum	23.44%	17.43%	13.78%
Confidence Level(95.0%)	2.06%	1.66%	1.40%

5.3. Dependant Faults Model Formulation

The previous section presented the results and conclusions for a preliminary study using the model formulated in Equation 4.3 to predict the yield of PCBs. Recall that one of the main assumptions of this model was to consider the formation of faults on a board independent of faults already on the board. This section will look at a model in which faults are assumed to be dependent on past faults. A comparison of the results for both models will be presented in the next section (Section 5.4.)

As was mentioned in Section 2.1, C. Stapper worked on developing simulation programs to model the frequency distribution of chip formation using the negative binomial distribution. In a paper presented in 1983, he mathematically showed the

validity of using the negative binomial distribution for modeling the yield of ICs [Stapper,1983]. The main assumption of this model *(Stapper's Assumption)* was:

The probability that a fault of type f_i occurs is independent of time and increases linearly with respect to the number of faults of type f_i that have already occurred.

Under this assumption faults are not considered independent as in the Poisson Model, but rather they are dependent on past fault history. This model does assume independence among the different component-type faults (Assumption #3).

The general form of the negative binomial yield model, can be derived following the same procedure used for the Poisson model derivation. This derivation is presented in Appendix C.

Stapper's Assumption is based on the likelihood that a fault for component type *i* (f_i) will occur given its fault history. The resulting model "clusters" faults by assuming that if one process has generated a faulty board, the likelihood of a fault occurring the next time the process takes place is greater than if no fault had occurred. The symbol α_i represents the clustering taking place for each process *i*.

The result of Stapper's analysis showed that the negative binomial yield model can be expressed as:

$$Y = \prod_{i=1}^{N} (1 + \frac{\lambda_i}{\alpha_i})^{-\alpha_i} \qquad (5.1)$$

where α_i is a clustering parameter for component-type *i*.

Stapper's negative binomial model appears difficult to understand intuitively since it accounts for no restrictions on the parameter's value. It therefore fails to meet the traditional interpretation of the negative-binomial distribution as the probability of occurrence for the "*k*th success".

Rogers [Rogers, 1974], explains that this form of the negative binomial distribution (referred to also as the *generalized negative binomial*), accounts for a behavior in which clustered dispersion is observed when doing quadrant analysis (breaking the study region into predefined equally sized areas and studying each area independently). Rogers compares this distribution with the Poisson distribution (random spatial dispersion) and the binomial distribution (regular spatial dispersion).

For the case in which α_i is unknown an optimization problem can be formulated in order to determine these values. As shown in Figure 5.2, this optimization model is similar to that presented for the Poisson model. The primary difference is that the α_i values are optimized instead of the p_i values (next page).

Given

NB = total number of different boards
j = board number, j = 1, ..., NB
y_j = actual yield for board j
N = total number of component types
$n_{j,i}$ = number of components of type i for board j

Find

α_i = clustering parameter for components of type i

Minimize

$$\sum_{j=1}^{NB} \left| \frac{y_j - \prod_{i=1}^{X} (1 + \frac{N_{j,i}}{\alpha_i})^{-\alpha_i}}{y_j} \right|$$

Subject to

$$\alpha_i \geq 0$$

Figure 5.2. Optimization Problem for Negative Binomial Yield Model

5.4. Comparison of Results for the Preliminary Case Study

The three same scenarios presented for the Poisson model were considered using the negative binomial model. This section presents the results for the preliminary case study using the negative binomial model and compares them with the results for the Poisson model. The results from negative binomial model were determined using the same search methods used for the Poisson model. Finally, conclusions are presented regarding the underlying fault formation assumptions made for each of these models.

The solution to the negative binomial optimization problem (Figure 5.2) for the single-variable model (Scenario #1) showed that on average this model predicts the yield of future board designs within 9.34% of the actual yield. With a standard deviation of 11.27%, a maximum difference with respect to the actual yield of 45.07%, and a 95% confidence level of 4.21%.

For the two-variable model (Scenario #2) the estimates were on average within 7.34% of the actual yield. The standard deviation of this result was determined to be 10.08%, with a with a maximum difference with respect to the actual yield of 39.41% and a 95% confidence level of 4.76%. These results showed very little improvement with respect to the single-variable case.

Finally, for the multi-variable case (Scenario #3) the optimization model (solved using genetic algorithms) showed that the predictions were on average within 4.82% of the actual yield. The standard deviation of this result was determined to be 8.08%, with a with a maximum difference with respect to the actual yield of 29.91% and a 95% confidence level of 3.02%. From these results it can be concluded that the addition of more design parameters improves the effectiveness of the model considerably.

A summary of the results from the three negative binomial scenarios is presented in Table 5.4. Appendix B also contains all the result tables for the thirty boards for the three negative binomial scenarios considered.

Each of the negative binomial scenario's results must be compared to those of the Poisson model in order to understand the significance of the results. These results as well as those for the Poisson model are summarized in Table 5.4. They include additional descriptive statistical information for comparison.

Table 5.4. Case Study Results for the Poisson and the Negative Binomial

	Scenario #1		Scenario #2		Scenario #3	
	Poisson	*Neg. Bin.*	*Poisson*	*Neg. Bin.*	*Poisson*	*Neg. Bin.*
N	1	1	2	2	14	14
Mean	5.73%	9.34%	3.76%	7.34%	2.91%	4.82%
Standard Error	1.01%	2.06%	0.81%	1.84%	0.69%	1.48%
Median	4.09%	6.80%	2.01%	4.32%	1.39%	1.17%
Standard Deviation	5.51%	11.27%	4.45%	10.08%	3.75%	8.08%
Sample Variance	0.30%	1.27%	0.20%	1.02%	0.14%	0.65%
Range	23.44%	45.07%	17.43%	39.41%	13.78%	29.91%
Minimum	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Maximum	23.44%	45.07%	17.43%	39.41%	13.78%	29.91%
Sum	171.81%	280.11%	112.92%	220.29%	87.17%	144.63%
Count	30	30	30	30	30	30
Largest(1)	23.44%	45.07%	17.43%	39.41%	13.78%	29.91%
Smallest(1)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Confidence Level(95.0%)	2.06%	4.21%	1.66%	3.76%	1.40%	3.02%

The results shown in Table 5.4 clearly show that for the preliminary case study, the Poisson model generates better yield estimates for the three scenarios considered. For the single-variable scenario the Poisson model yield estimates are, on average, close to 40% better than the ones form the negative binomial model. For the two-variable scenario, the Poisson results are better by 50%. Finally, for the multi-variable scenario the difference between the two is also close to 40%.

One of the main objectives in performing this case study was to determine the most valid assumption regarding the formation of faults in the board. Since the Poisson model generates considerably better results, it appears that the fault independence assumption is better than the dependant assumption. Consequently, it also strongly suggests that faults do not appear to cluster.

CHAPTER VI

YIELD MODEL REFORMULATION

This chapter presents two steps that will define the implementation of the yield model. The first step reformulates the model in order to simplify it. The second step considers additional constraints that may want to be considered when estimating yield. Chapter VII provides a detailed mathematical analysis of the model and the results. Finally, Chapter VIII presents a case study to show the overall process of implementing the yield model on actual production line data and summarizes the results.

6.1. Model Reformulation

In Chapter IV, a methodology was presented in which an optimization problem was formulated to minimize the difference between the actual yield and the predicted yield for some previous yield history. The Poisson model provided more reliable yield estimates than those of the negative binomial model, hence this will be the model used. This section examines the main difficulties of this optimization problem based on the Poisson model and reformulates it in order to overcome some of them.

The Poisson optimization model presented in Figure 4.2 is a constrained non-linear optimization problem. It is non-linear because: (1) the objective function takes the absolute value of the average difference between expected and actual yields, and (2) yield is estimated using exponential coefficients. The non-linear nature of this problem makes it extremely difficult to solve. Different search methods can be used to solve this type of problem. Some of the most common are: simulated annealing, genetic algorithms, and generalized hill-climbing algorithms [Bazaara, 1993].

The preliminary case study presented in Chapter V showed results for the single-variable, two-variable, and multi-variable scenarios studied. The results for the multi-variable scenario were obtained using a genetic algorithm software tool, Evolver 4.0 (Palisade Corp.), which is an add-in for Excel. This approach was chosen because it was convenient to implement although other search methods could have been used. As was mentioned in Chapter V the drawbacks of this approach are that: the solution is not guaranteed to be optimal (heuristic), and the computational time for arriving to the solutions can be extremely long. For example, for the multi-variable case the model was stopped after 17 hours. This might not be favorable for a modeling application that needs to become more efficient at predicting yields by constantly reformulating the problem as more data becomes available.

Recall that the objective of the optimization model is to provide an estimate for the component-type fault spectrum. This is achieved by minimizing the difference between actual and predicted yields. Under the assumption that faults are independent (Poisson model), the model can be reformulated to make the search process more efficient following these steps:

1. *Take the natural log of both the actual and the expected yields.*

Let c_j represents the negative log of the actual yield for board j such that:

$$c_j = -LN(y_j) .$$

The result of taking the log of the predicted yield (Equation 4.3) is

$$\ln(Y) = \ln(e^{-\sum_{i=1}^{N}\lambda_i}) = -\sum_{i=1}^{N}\lambda_i .$$

2. *Square the difference between actual and expected yields.*

The purpose of taking the absolute value in the objective function of the original optimization problem is to avoid negative differences. This can also be done by squaring the function.

Incorporating these changes into the original problem transforms the objective function. In particular, instead of minimizing the difference between actual and predicted yields the function minimizes the square of the difference between the log of the actual yield and the sum of the fault rates over all the different component-types considered. The optimization problem is reformulated as shown in Figure 6.1.

Given

NB = total number of different boards

j = board number, j = 1, ..., NB

y_j = actual yield for board j

N = total number of component types

$n_{j,i}$ = number of components of type i for board j

$c_j = -LN(y_j)$

Find

p_i = fault probability for components of type i

Minimize

$$\sum_{j=1}^{NB} (c_j - \sum_{i=1}^{N} p_i n_{j,i})^2$$

Subject to

$$0 \le p_i \le 1$$

Figure 6.1. Reformulated Poisson Optimization Problem

The advantage of using this problem is that neither exponential terms nor absolute values are present in the objective function making it considerably easier to solve.

6.2. Additional Constraints

The ability of the yield model to adjust to desired specifications is extremely important. In particular, when estimating yield of production lines, some desired constraints can be included depending on the application. For example, it might be of greater importance not to under-estimate than to over-estimate the manufacturing yield of

a particular product. This section considers these two additional optional constraints and reformulates the optimization problem to account for each.

6.2.1. Over-Estimation Constraint

The first additional constraint considered accounts for cases when it is highly undesirable to over-estimate manufacturing yield. This constraint is of particular interest for situations in which forecasts of production are based on yield estimates and it is extremely costly to fail to meet the demand. To avoid these high loss costs, the manufacturer may prefer to take a conservative approach by constraining the expected yield model.

Suppose that the difference between the predicted and the actual yields is recorded and the sum over all of the boards considered is minimized. Note that for the case in which yield is never over-estimated, the difference between actual and predicted yield must always be a positive number. Using the optimization problem presented in the previous section (Figure 6.1) a similar analysis can be performed. Recall that for this model the difference between the log of the actual yield and the sum of the fault rates is minimized. The quantity a_j can represent this difference for each board design (j) considered. In addition, it is desired that this variable will always be non-negative.

The addition of this constraint transforms the non-linear optimization problem presented in Figure 6.1 into a linear programming (LP) optimization problem. The resulting LP is presented in Figure 6.2.

Given

NB = total number of different boards
j = board number, j = 1, …, NB
y_j = actual yield for board j
c_j = log of actual yield, ln(y_j)
N = total number of component types
$n_{j,i}$ = number of components of type i for board j
a_j = difference in between expected and actual yields for board j

Find

p_i = fault probability for components of type i

Minimize

$$\sum_{j=1}^{NB} a_j$$

Subject to

$$p_i \leq 1$$
$$a_j, p_i \geq 0$$

Such that

$$c_j - \sum_{i=1}^{N} p_i n_{j,i} = a_j \quad \forall j=1,2,...,NB$$

Figure 6.2. Over-Estimation Linear Program

6.2.2. Under-Estimation Constraint

Another situation that might be of interest is to add a constraint so as to never under-estimate the actual yield. This can be done following the same procedure as that explained in the previous section. The only difference is that the difference constraints (a_j *values*) will be restricted to negative values.

Although intuitively less applicable than the over-estimation constraint, the under-estimation constraint could be useful for cases in which excess inventories are highly undesirable. This might be the case, for example, in the production of items with special storage requirements that in the event of excess production need to be discarded. In this situation the manufacturer might consider adding the under-estimation constraint in order to plan storage requirements appropriately so as to never discard produced items due to lack of storage space.

Given

NB = total number of different boards
j = board number, j = 1, …, NB
y_j = actual yield for board j
c_j = log of actual yield, $\ln(y_j)$
N = total number of component types
$n_{j,i}$ = number of components of type i for board j
a_j = difference in between expected and actual yields for board j

Find

p_i = fault probability for components of type i

Minimize

$$\sum_{j=1}^{NB} a_j$$

Subject to

$$p_i \leq 1$$
$$a_j, p_i \geq 0$$

Such that

$$\sum_{i=1}^{N} p_i n_{j,i} - c_j = a_j \quad \forall j=1,2,...,NB$$

Figure 6.3. Under-Estimation Linear Program

CHAPTER VII

YIELD MODEL ANALYSIS

This chapter provides a detailed mathematical analysis of the yield model using the production line data from the preliminary case study. The performance of the yield model is measured by performing four different analyses using the reformulated optimization model (Figure 6.1). First, the solution times between the original model and the reformulated model are compared. Second, the performances of the constrained models (with under and over estimation constraints) are compared to the results from the reformulated model. Next, a sensitivity analysis is performed in order to determine the effect that each component type had on the manufacturing yield. Finally, the performance of different board designs is determined making use of the results from the previous analysis.

7.1 Solution Times for Preliminary Case Study

Solution times for the preliminary case study are presented in this section. Of particular interest is the comparison of the solution times for the original and reformulated models. Also of interest is the comparison of the solution times required for various search methods and/or software packages. The results presented in this section are based on the multi-variable scenario, since it is the only one that requires a considerable amount of time to be solved.

As mentioned in Section 5.2, the genetic algorithm approach used to solve the original optimization yield model was extremely time consuming (~17hrs). The reformulated optimization model is considerably faster (~15 min) using the same solution

apptoach. The results from this model are on average within 3.06% of the actual yield with a standard deviation of 3.27%. The maximum difference between the expected and the actual yield is 11.56% (see Table 7.1).

An exact comparison of these solution times is difficult to assert due to the heuristic nature of the algorithm. However, the drastic difference in times is enough to show clearly that solving the reformulated problem is considerably faster and provides very close (if not better) results than those obtained from the original model.

Other search methods were tested using various software packages to show the performance of the reformulated model. Quasi Newton search methods and conjugate gradient search methods were implemented using the built in Excel Solver developed by Frontline System. Mathematica was used to solve the model using the built in FindMinimum command which searches for an optimal solution using a steepest descent approach. Finally, Evolutionary by Frontline Systems, a genetic algorithm tool, was used to compare its performance with the original genetic algorithm tool used, Evolver by Palisade Corp. The results from these search methods as well as the solution times are summarized in Table 7.1.

Table 7.1. Results and Solution Times from Different Search Methods

Tool Name	Software Company	Search Method	Ave.Diff. w.r.t. Actual Yield	St. Dev.	Max. Diff. w.r.t. Actual Yield	Solution Time
Evolver*	Palisade Corp.	Genetic Algorithms	2.91%	3.75%	13.78%	~17 hr
Evolver	Palisade Corp.	Genetic Algorithms	3.06%	3.27%	11.56%	~15 min
Solver	Frontline Systems	Q-Newton Search	6.95%	4.30%	18.50%	~35 sec
Solver	Frontline Systems	Conjugate Gradient	95.62%	14.76%	100.00%	~2 min
Evolutionary	Frontline Systems	Genetic Algorithms	3.04%	3.26%	11.56%	~20 min
Mathematica	Wolfram Research	Steepest Descent	3.08%	3.25%	11.55%	~30min

* Original Model

The results presented in Table 7.1 were all generated using the same initial random starting solution. For this reason, some search methods such as the conjugate gradient failed to perform well. Quasi Newton search methods and conjugate search methods tend to get trapped in their search processes once they reach certain levels. In addition, these methods exhibit extreme sensitivity to initial conditions [Bazaraa, 1993].

A combination approach was suggested in order to overcome this problem. In particular, the proposed algorithm starts at any given random solution and performs three consecutive quasi Newton searches. This implies that the solution for the first search becomes the starting point for the next one and so on. Subsequently, the algorithm performs two conjugate search methods. Finally, a last quasi Newton search is performed. The results from this algorithm are summarized in Table 7.2.

Table 7.2. Quasi Newton & Conjugate Gradient Algorithm Result

Tool Name	Software Company	Search Method	Ave.Diff. w.r.t. Actual Yield	St. Dev.	Max. Diff. w.r.t. Actual Yield	Solution Time
Solver	Frontline Systems	Q-Newton & Conjugate Gradient	3.05%	3.26%	11.59%	~60 sec

By comparing the results presented in Tables 7.1 and 7.2, it can be seen that the algorithm involving multiple quasi Newton & conjugate gradient searches clearly performs very well. This algorithm offers considerably reductions in solution times. Another key attribute of this algorithm is that it can be easily coded using built-in functions of Microsoft Excel, minimizing the amount of coding and file transfer (code adaptation) required.

Although this algorithm apparently does not perform as well as the genetic algorithm approach (2.91% vs. 3.05% average difference with respect to the actual yield), further analysis of the results show that the difference between the two is questionable. In fact, the considerable reduction in standard deviation (3.76% vs. 3.26%) and in maximum difference (13.78% vs. 11.56%) strongly suggest that the proposed algorithm is as effective in determining reliable estimates as the genetic algorithm approach. The standard deviation and the maximum difference are reduced because the revised optimization problem squares the objective function instead of taking its absolute value.

In general, it can be concluded that the best algorithm for the desired application is the combined quasi Newton & conjugate gradient algorithm. The results from this algorithm offer comparable solutions to those obtained using genetic algorithm approaches in considerably time.

7.2 Constraint Model Results for Preliminary Case Study

The previous section highlighted the difference in computational time using different search methods and software packages. This section will present the results obtained from the linear optimization problems presented in Section 6.2 (yield over-estimation and under-estimation). The motivation for analyzing these problems is that the computational time for solving linear problems is negligible compared to the time required to solve non-linear ones. The objective of this section is to draw conclusions regarding the applicability of these constrained problems.

The linear optimization problems were solved using the built in Solver (Frontline Systems) for Microsoft Excel. Table 7.3 summarizes the results from both the under and the over estimation yield models.

Table 7.3. Results from the Constrained Yield Models

Tool Name	Software Company	Search Method	Ave. Diff. w.r.t. Actual Yield	St. Dev.	Max. Diff. w.r.t. Actual Yield	Solution Time
Solver	Frontline Systems	Linear Program Under Estimate	10.43%	16.01%	61.31%	~5 sec
Solver	Frontline Systems	Linear Program Over Estimate	6.61%	9.50%	36.23%	~5 sec

From the results presented above it can be concluded that using the additional constraints will imply potential benefits and pitfalls to the yield model. As expected the computational time to reach the solution was considerably lower than that for the non-linear model solution times presented in the previous section. The model was also considerably easier to implement due to its linear nature. However, a closer look at the

results shows that the estimates are not as good as those from the non-linear model. The linear problem only searches the solution space from one-side since the model is designed to either under or over estimate, but not both at the same time. This makes the results less accurate.

The applicability of the constrained model lies in the quality of the data that is available, the application for which it is intended, and the desired accuracy of the yield estimates. If the input data is not very dispersed (relatively small actual yield range), the model will be able to perform better since it is more likely to find neighboring solutions that fit the constraints. As mentioned in Section 6.2 the constrained models depend on the application for which they are intended. In general, if the yield estimates need not to be very accurate (for example for rough capacity plans) and results need to be obtain in a very fast way, the constrained yield model may be an ideal choice.

7.3 Sensitivity Analysis of Component Types for Preliminary Case Study

This section presents a sensitivity analysis study of the effect that each component type has on the manufacturing yield of a given board design. When a new board design is scheduled for production, engineers not only want to predict the yield but they also want to know how they can improve it. This sensitivity analysis provides information to assist in improving the yield.

The results for the sensitivity analysis were obtained using the combined quasi Newton and conjugate gradient algorithm presented in Section 7.3 (implemented using Microsoft's Visual Basic for Applications). Figure 7.1 presents a graph that shows the effect that each component type has on yield (see Table 5.1 for component descriptions).

The graph shows an increment of 1 to 20 for each component type assuming that all other components remain constant. Each increment consists of adding a component of a given type while the rest of the components in the board remain the same. The purpose of considering individual increments is to measures the effect that adding one component has on yield.

From the results presented in Figure 7.1 two component types (*JD* and *nsmthd*) clearly appear to be affecting manufacturing yield considerably. *JD* stands for the number of plastic leaded chip carriers (PLCCs) and *nsmthd* stands for non-standard manual through devices. The graph shows that an increment of 20 *JD* components can decrease manufacturing yield by as much as 8.9%. The same increment of *nsmthd* components will lower the yield by 5.5%. For comparison, an increment of 20 dual in line packages (*DIPs*) will decrease yield by only 0.02%.

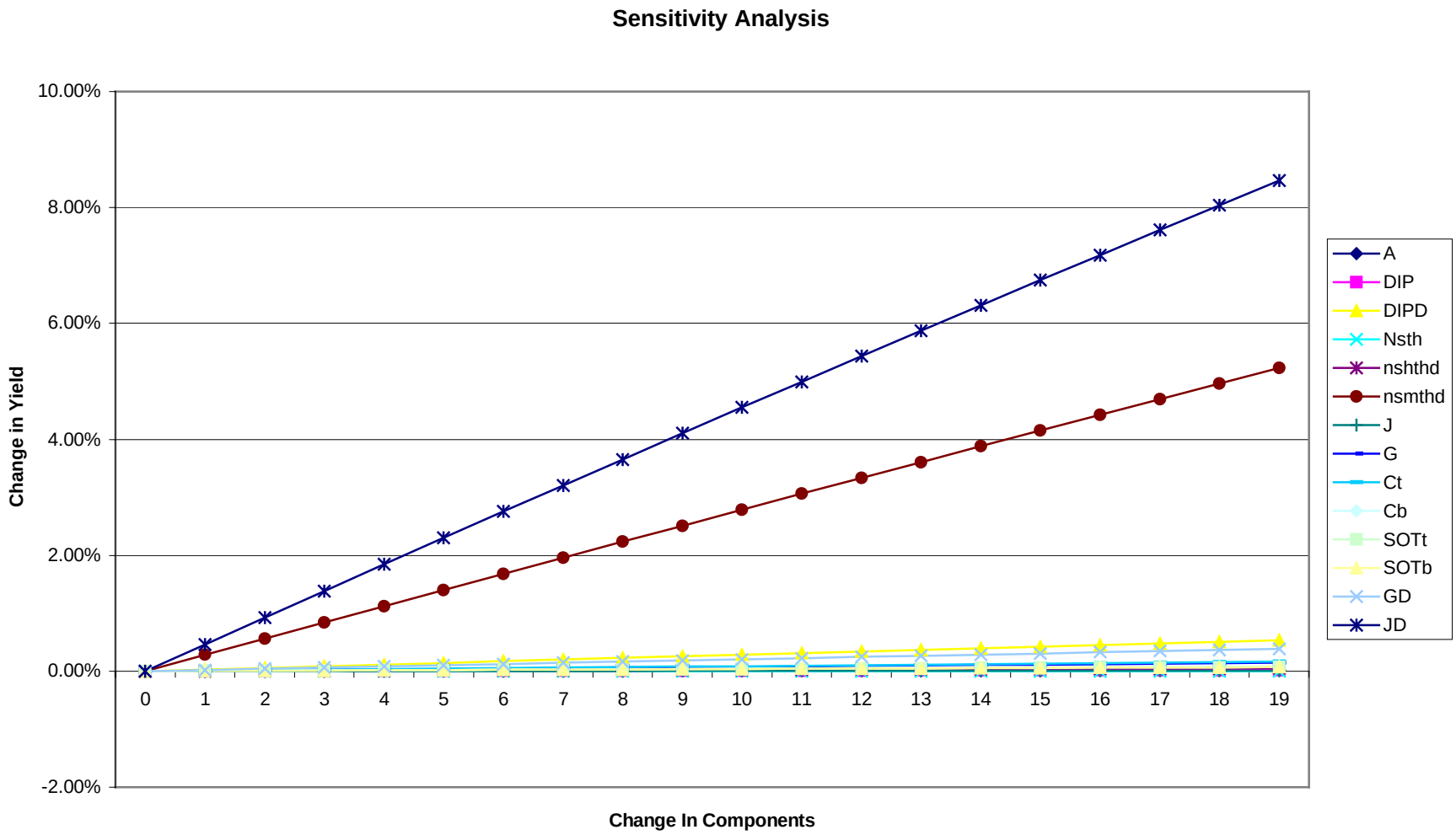

Figure 7.1. Sensitivity Analysis of Component-Types

The results from the sensitivity analysis study provide a direct trade-off analysis between component types and their effect on yield. Engineers can utilize this information to determine which board design best suits the performance characteristics of their given production lines. The following section presents a method in which this information can be used to evaluate new board designs.

7.4 Board Design Evaluation for Preliminary Case Study

The main objective of this research is to develop a tool that will aid manufacturers of electronic products in determining reliable estimates of production capabilities as they introduce new PCBs onto their production lines. This section summarizes how this can be done by analyzing the expected performance of a new board design by making use of the results presented in the previous sections of this chapter.

Assume that a given board design requires 50 components of each of the fourteen component/lead families. Using the fault probability spectrum from Section 7.1 and Equation 4.3, the yield of this design can be estimated to be 66.03%. A series of four changes to this initial design were evaluated and their effect on yield is presented in Table 7.4.

Table 7.4. Evaluation of Different Board Designs

Board Design	A	DIP	DIPD	Nsth	nshthd	nsmthd	J	G	Ct	Cb	SOTt	SOTb	GD	JD	Yield
1	50	50	50	50	50	50	50	50	50	50	50	50	50	50	66.03%
2	50	50	50	50	50	**20**	50	50	50	50	50	50	50	**40**	75.30%
3	**75**	50	50	50	50	50	50	**0**	50	50	50	50	50	50	66.27%
4	88	90	53	77	84	**0**	54	74	60	86	57	52	93	**0**	94.47%
5	25	71	72	36	31	**75**	52	90	0	55	43	75	83	**80**	52.86%

1. The first change considered was intended to see the effect that reducing the number of *nsmthd* and *JD* components would have on yield. The number of *nsmthd*s was reduced by 30 and the number of *JD*s was reduced by 10. This change had a significant impact on the predicted yield changing it to 75.30% (a 14% increase with respect to the original design).
2. The second design change analyzed was to increase the number of axial leaded components (*A*s) from 50 to 75, and to eliminate gull wing small outline integrated circuits SOIC (*Gs*). This change did not have a significant impact on the predicted yield, changing it by merely 0.36%.
3. The third design change randomly increased the number of components for all component types except for *nsmthd*s and *JD*s which were eliminated. The effect that this change had on the predicted yield was very significant increasing yield by as much as 43.1%.
4. Finally, the last design change represented a random board design. Since both the total number of *nsmthd*s and *JD*s increased for this random design, the predicted yield for this was lower than that for the original board design. The model predicts a yield loss of 20% with respect to the original board design for this random design.

These four changes are presented to show the applicability of the yield model when analyzing the performance of future board designs on current production lines. More detailed what-if scenarios can be considered, and more specific design constraints can be included. However, the point is to illustrate that yield can be predicted based on

the past performance of different board designs. Furthermore, design suggestions can be made in order to account for manufacturability issues at the design level. This is an area for future research and if studied in more depth could lead to the refinement of concepts such as design for manufacturability (DFM).

CHAPTER VIII

CASE STUDY DEVELOPMENT AND RESULTS

The past seven chapters have presented the development, validation, refinement, reformulation, implementation and analysis of a model to predict the yield for new PCBs. This analysis was done based on production line data available from the literature. The following chapter presents a new case study that was designed to test the proposed yield model on new current production line data.

The chapter is divided into four main sections. The first section describes the selection of a suitable production line and the data gathering process. The following section describes the component families considered in the case study. The results from the yield model implementation and the sensitivity analysis of the component families are presented in the last two sections.

8.1. Production Line Selection and Data Gathering

In order to select a suitable production line to perform this case study, a questionnaire was developed and distributed in a multinational electronic assembly firm that sponsored this research. The content of the questionnaire is shown in Figure 8.1. The questionnaire was targeted to production line engineers that could benefit from the potential results of the study. The corporate sponsor assisted in the distribution (via electronic format) of the questionnaire to the appropriate facilities around the world.

A total of twelve different facilities from around the world responded to the questionnaire. After careful selection, a European facility was selected as the most appropriate for conducting the case study due to the availability of the required data.

(Contract agreements do not permit the name of the facility to be disclosed in this document).

Manufacturing Design Case Study

QUESTIONNAIRE:

Name and location of production line? _____

Have you assembled more than one Board Designs using one production line?
__ Yes ____No

If Yes, how many different Board Designs do you estimate the production line has assembled?

_____ Less than 5 different board designs
_____ Between 5 and 10 different board designs
_____ Between 10 and 20 different board designs
_____ More than 20

If Yes, do you know the component characteristics of the Board Designs produced (i.e. component families on each board and the quantity of each family)?

____Yes ___No

*If Yes, do you have the actual production yields** for each of the Board Designs produced?*
____Yes ____No

If Yes, please fill the data table in the following page (feel free to adjust the table to fit your data).

Any additional information that you can provide regarding the board designs, production line, or component families will be of great value.

Virginia Tech Research Team

**Yield is defined as the ratio of good boards produced over the total number of boards produced.

Figure 8.1. Case Study Questionnaire

The facility chosen had produced over 20 different board designs and had collected all the required data to perform the case study. Data collection at this facility was done using a Microsoft Access database. After further contact with a production engineer from the facility, data from a total of two Fuji surface mount assembly lines was provided to perform the case study.

8.2. Component Families Considered for the Case Study

After gathering the data, the next task was to group the components into families. This was done by considering the electronic package type of each component. Those components with similar electronic package types were grouped into one family. Electronic packages vary in size, number of pins, mounting type (through hole or surface mount), and material.

A comprehensive classification of electronic packages is available at the SMA web page from Harvard University. The electronic link to this site is: http://sma-www.harvard.edu/private/file_view/parts_lib/index.html . This site also provides visual images of the package devices. Additional electronic package classifications can be found in Coombs [Coombs, 1988] and Rowland [Rowland, 1993].

Due to an agreement with our corporate sponsor the name of the actual electronic packages will not be presented in this report. A total of 11 different component families were utilized for the case study. Data representing each package type is referred to as fam-1 through fam-11. Two examples of these families (package types) include small outline transistors (SOTs) and shrink small outline packages (SSOPs).

Although data from two different production lines was provided by the European facility, a third line was also included in the analysis. This line corresponds to the aggregate of the other two lines (it could be thought of as a particular production group within a facility). The distribution of component families per production line is presented in Figure 8.2. The figure shows the percentage of components in each family with respect to the total number of components placed by each production line for the board designs considered. Note that PL-1 (production line #1) and PL-2 (production line #2)

represent the data from the individual production lines, whereas PL-3 (production line #3) represents the data from the production group compromised by PL-1 and PL-2.

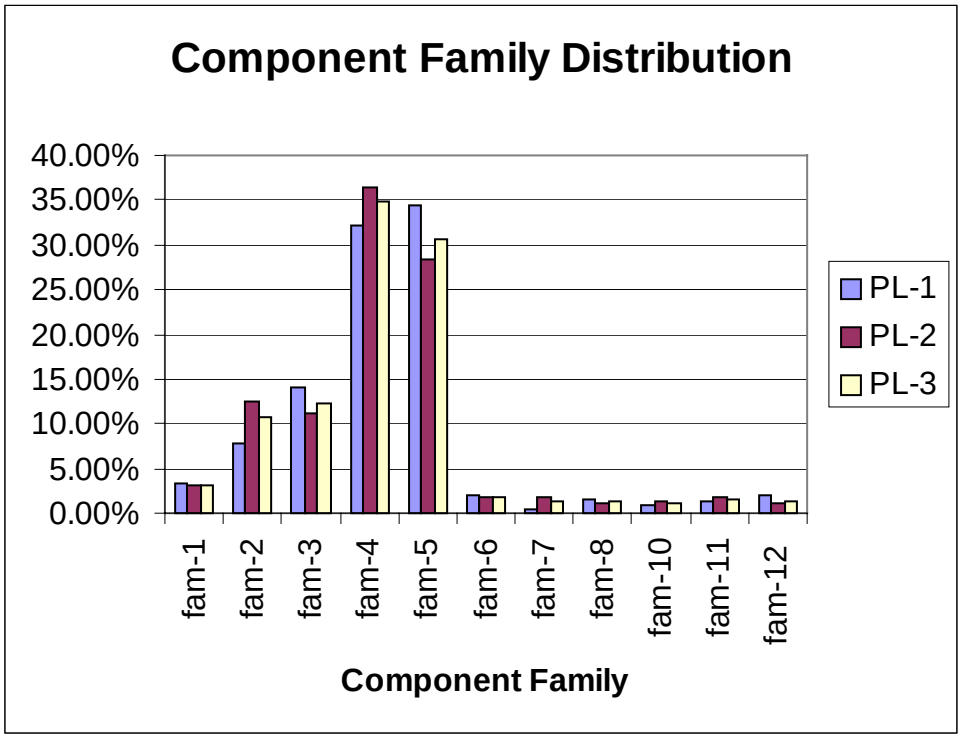

Figure 8.2. Component Family Distribution

From the above figure it can be seen that:

- Component families fam-4 and fam-5 account for approximately 60% of the total number of components placed.
- Approximately 25% of the components are split between fam-1, fam-2, and fam-3.
- Approximately 15% is divided evenly among families 6 through 12.

8.3. Case Study Implementation and Results

From the analysis done in the previous seven chapters it was determined that the best model for predicting yield is the reformulated non-linear optimization problem (Figure 6.1). The method selected for solving the model was the combined quasi-Newton & conjugate gradient search algorithm (Section 7.3). Data from PL-1 consisted of 12 different board designs with actual yields ranging from 91.90% to 100%. Data from PL-2 consisted of 13 different boards with yields ranging from 83.40% to 99.60%. Data from PL-3 consisted of 25 different board designs and ranged from 83.40% to 100%. Each board design considered in each line was assigned a name that described which line it was produced in. For example, board design number 5 for PL-1 was numbered BD1-5.

Tables 8.1, 8.2 and 8.3 present the following information for each production line:

- board designs,
- total number of components per component family,
- total number of components per board design,
- actual yield per board design,
- predicted yield per board design, and the
- difference between the predicted and actual yields per board design.

Figures 8.3, 8.4 and 8.5 illustrate the average difference between the predicted and the actual yield for each production line considered. Statistical data regarding these average differences is also provided with these figures. The figures present the data arranged from the minimum difference to the maximum difference with respect to the actual yield.

Finally, Figure 8.6 illustrates the fault spectrum for all three production lines is shown. These values were optimized using the quasi-Newton & conjugate gradient algorithm, and were also the values used to estimate the yields using the Poisson model in Tables 8.1, 8.2 and 8.3.

8.4. Sensitivity Analysis of Case Study Results

A sensitivity analysis study was conducted in order to determine the effect that each component family has on the manufacturing yield of a given board design. The results for the sensitivity analysis were obtained using the combined quasi Newton & conjugate gradient algorithm presented in Section 7.3 implemented using Microsoft's Visual Basic for Applications. Figures 8.7, 8.8 and 8.9 present graphs that show the effect that each component family (electronic package type) has on yield. The graph shows the effect on yield of increasing the number of a particular component type assuming that the number of all other components remains constant.

From the sensitivity analysis graphs it can be seen that two component families considerably affect yield more than the rest. For PL-1, the families that affect yield the most are fam-10 and fam-6. For PL-2, the families that affect yield the most are fam-10 and fam-7. Finally for PL-3, the families that affect yield the most are also fam-10 and fam-7.

Table 8.1. Summary Data and Results for PL-1

Board Design No.	Total No. of Components per Component Family											Total No. of Comp.	Actual Production Yield	Predicted Production Yield	Difference
	fam-1	fam-2	fam-3	fam-4	fam-5	fam-6	fam-7	fam-8	fam-9	fam-10	fam-11				
BD1-1	0	3	9	5	0	3	0	4	0	0	0	24	100.00%	**99.53%**	0.47%
BD1-2	1	3	13	16	1	2	0	1	0	0	0	37	99.00%	**99.68%**	0.69%
BD1-3	4	24	25	50	2	9	1	2	0	15	0	132	91.90%	**95.88%**	4.33%
BD1-4	0	12	24	32	8	4	2	0	0	12	0	94	98.10%	**97.19%**	0.93%
BD1-5	2	52	121	81	2	11	1	2	1	3	0	276	97.10%	**97.72%**	0.64%
BD1-6	2	52	121	79	2	11	1	2	1	3	0	274	98.30%	**97.72%**	0.59%
BD1-7	2	52	121	81	2	11	1	2	1	3	0	276	100.00%	**97.72%**	2.28%
BD1-8	14	5	0	43	13	2	2	2	1	14	0	96	100.00%	**97.10%**	2.90%
BD1-9	14	16	18	182	186	18	0	37	0	8	0	479	96.00%	**95.72%**	0.30%
BD1-10	46	21	0	274	756	6	4	0	20	0	34	1161	97.40%	**97.64%**	0.25%
BD1-11	16	3	108	189	0	2	2	3	9	0	22	354	97.20%	**98.90%**	1.75%
BD1-12	28	26	0	128	194	2	2	3	2	0	11	396	100.00%	**99.26%**	0.74%
BD1-13	3	44	0	130	206	0	2	2	2	0	10	399	99.30%	**99.65%**	0.35%
Sum of Comp. per Family	132	313	560	1290	1372	81	18	60	37	58	77			**Average**	**1.25%**
														Stdv	**1.24%**
	P(fam-1)	**P(fam-2)**	**P(fam-3)**	**P(fam-4)**	**P(fam-5)**	**P(fam-6)**	**P(fam-7)**	**P(fam-8)**	**P(fam-9)**	**P(fam-10)**	**P(fam-11)**			**Max**	**4.33%**
Optimized Component Type Fault Spectrum	1.74E-05	1.52E-07	4.13E-09	1.00E-07	1.46E-06	1.59E-03	4.80E-08	1.36E-06	1.29E-04	1.85E-03	2.91E-04			**Yield Range**	**8.10%**

Table 8.2. Summary Data and Results for PL-2

Board Design No.	Total No. of Components per Component Family											Total No. of Comp.	Actual Production Yield	Predicted Production Yield	Difference
	fam-1	fam-2	fam-3	fam-4	fam-5	fam-6	fam-7	fam-8	fam-9	fam-10	fam-11				
BD2-1	16	109	204	379	7	7	7	4	2	5	2	742	89.50%	**88.29%**	1.35%
BD2-2	1	3	13	16	1	2	0	1	0	0	0	37	98.80%	**99.31%**	0.51%
BD2-3	12	64	198	146	4	14	1	0	0	14	4	457	85.50%	**87.16%**	1.94%
BD2-4	3	57	121	79	2	11	1	2	1	3	0	280	95.90%	**93.16%**	2.86%
BD2-5	16	22	18	182	186	18	0	37	0	8	0	487	99.60%	**96.09%**	3.53%
BD2-6	16	69	0	44	30	3	11	0	7	1	0	181	97.20%	**97.23%**	0.03%
BD2-7	14	124	0	372	227	5	21	7	24	29	10	833	92.50%	**88.00%**	4.86%
BD2-8	10	115	0	258	125	12	15	9	18	34	0	596	84.10%	**87.60%**	4.16%
BD2-9	47	30	0	274	756	6	12	0	20	0	34	1179	98.50%	**97.12%**	1.40%
BD2-10	25	94	141	256	19	26	24	6	1	11	0	603	83.40%	**85.25%**	2.22%
BD2-11	29	37	0	136	206	2	9	3	2	0	11	435	94.90%	**97.97%**	3.24%
BD2-12	4	52	0	130	206	0	9	2	2	0	10	415	99.20%	**98.08%**	1.13%
Sum of Comp. per Family	193	776	695	2272	1769	106	110	71	77	105	71			**Average**	**2.27%**
														Stdv	**1.48%**
	P(fam-1)	**P(fam-2)**	**P(fam-3)**	**P(fam-4)**	**P(fam-5)**	**P(fam-6)**	**P(fam-7)**	**P(fam-8)**	**P(fam-9)**	**P(fam-10)**	**P(fam-11)**			**Max**	**4.86%**
Optimized Component Type Fault Spectrum	1.04E-06	1.18E-08	4.50E-04	4.24E-08	4.86E-08	5.41E-04	2.15E-03	2.86E-08	3.64E-08	2.75E-03	3.53E-08			**Yield Range**	**16.20%**

Table 8.3. Summary Data and Results for PL-3

Board Design No.	Total No. of Components per Component Family											Total No. of Comp.	Actual Production Yield	Predicted Production Yield	Difference
	fam-1	fam-2	fam-3	fam-4	fam-5	fam-6	fam-7	fam-8	fam-9	fam-10	fam-11				
BD2-1	16	109	204	379	7	7	7	4	2	5	2	742	89.50%	**90.67%**	1.31%
BD2-2	1	3	13	16	1	2	0	1	0	0	0	37	98.80%	**99.48%**	0.69%
BD2-3	12	64	198	146	4	14	1	0	0	14	4	457	85.50%	**89.88%**	5.13%
BD2-4	3	57	121	79	2	11	1	2	1	3	0	280	95.90%	**94.76%**	1.18%
BD2-5	16	22	18	182	186	18	0	37	0	8	0	487	99.60%	**96.42%**	3.19%
BD2-6	16	69	0	44	30	3	11	0	7	1	0	181	97.20%	**96.71%**	0.51%
BD2-7	14	124	0	372	227	5	21	7	24	29	10	833	92.50%	**87.80%**	5.08%
BD2-8	10	115	0	258	125	12	15	9	18	34	0	596	84.10%	**87.74%**	4.33%
BD2-9	47	30	0	274	756	6	12	0	20	0	34	1179	98.50%	**96.44%**	2.10%
BD2-10	25	94	141	256	19	26	24	6	1	11	0	603	83.40%	**86.12%**	3.27%
BD2-11	29	37	0	136	206	2	9	3	2	0	11	435	94.90%	**97.48%**	2.72%
BD2-12	4	52	0	130	206	0	9	2	2	0	10	415	99.20%	**97.66%**	1.55%
BD1-1	0	3	9	5	0	3	0	4	0	0	0	24	100.00%	**99.55%**	0.45%
BD1-2	1	3	13	16	1	2	0	1	0	0	0	37	99.00%	**99.48%**	0.49%
BD1-3	4	24	25	50	2	9	1	2	0	15	0	132	91.90%	**94.85%**	3.21%
BD1-4	0	12	24	32	8	4	2	0	0	12	0	94	98.10%	**95.63%**	2.52%
BD1-5	2	52	121	81	2	11	1	2	1	3	0	276	97.10%	**94.77%**	2.40%
BD1-6	2	52	121	79	2	11	1	2	1	3	0	274	98.30%	**94.77%**	3.59%
BD1-7	2	52	121	81	2	11	1	2	1	3	0	276	100.00%	**94.77%**	5.23%
BD1-8	14	5	0	43	13	2	2	2	1	14	0	96	100.00%	**95.94%**	4.06%
BD1-9	14	16	18	182	186	18	0	37	0	8	0	479	96.00%	**96.43%**	0.45%
BD1-10	46	21	0	274	756	6	4	0	20	0	34	1161	97.40%	**98.48%**	1.11%
BD1-11	16	3	108	189	0	2	2	3	9	0	22	354	97.20%	**96.08%**	1.16%
BD1-12	28	26	0	128	194	2	2	3	2	0	11	396	100.00%	**99.28%**	0.72%
BD1-13	3	44	0	130	206	0	2	2	2	0	10	399	99.30%	**99.47%**	0.17%
Sum of Comp. per Family	325	1089	1255	3562	3141	187	128	131	114	163	148			**Average**	**2.26%**
														Stdv	**1.64%**
	P(fam-1)	**P(fam-2)**	**P(fam-3)**	**P(fam-4)**	**P(fam-5)**	**P(fam-6)**	**P(fam-7)**	**P(fam-8)**	**P(fam-9)**	**P(fam-10)**	**P(fam-11)**			**Max**	**5.23%**
Optimized Component Type Fault Spectrum	2.76E-05	2.42E-08	3.07E-04	4.58E-08	3.12E-08	5.90E-04	2.62E-03	3.65E-08	2.80E-08	2.48E-03	7.33E-08			**Yield Range**	**16.60%**

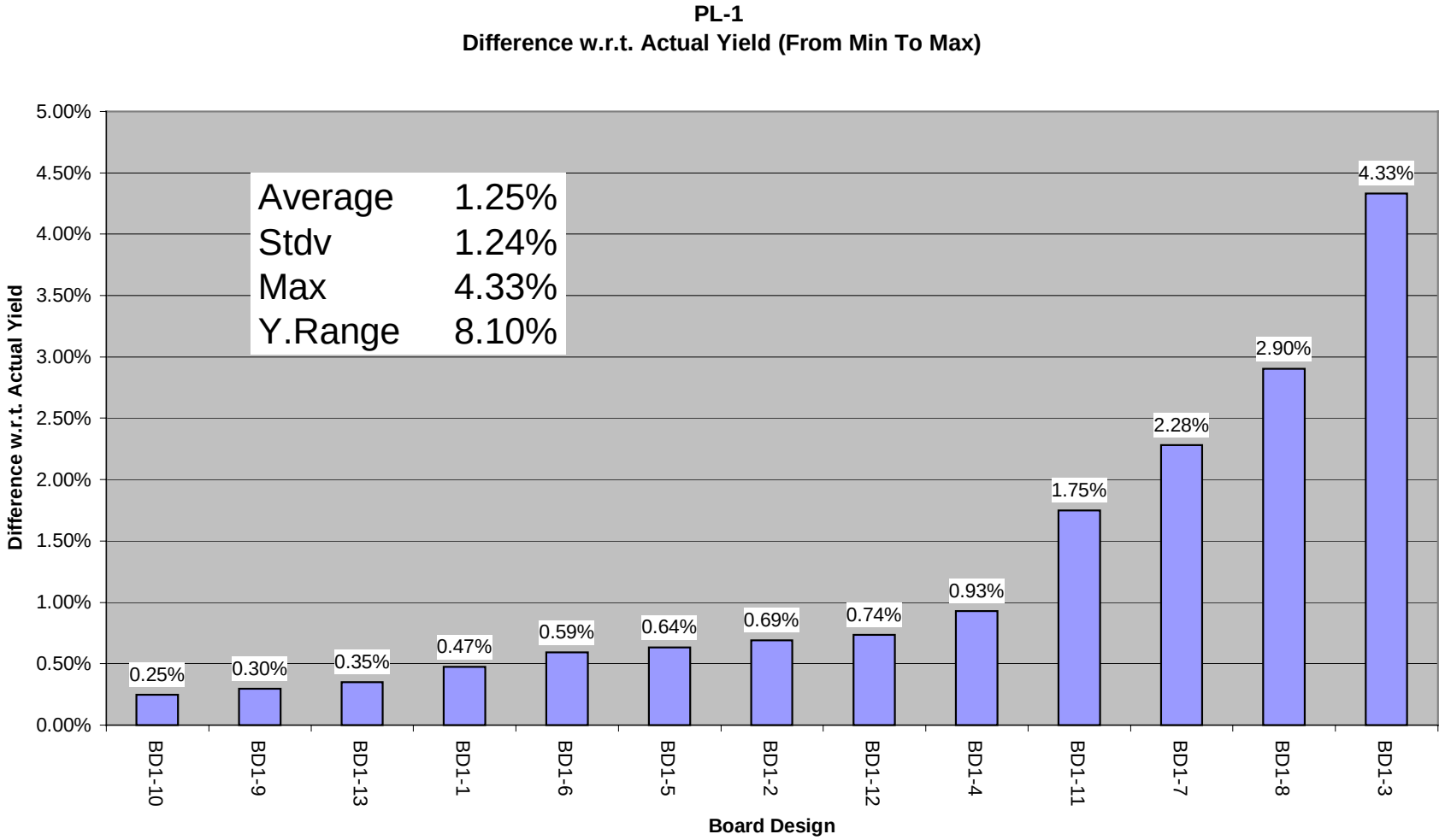

Figure 8.3. Average Difference With Respect to Actual Yield for PL-1

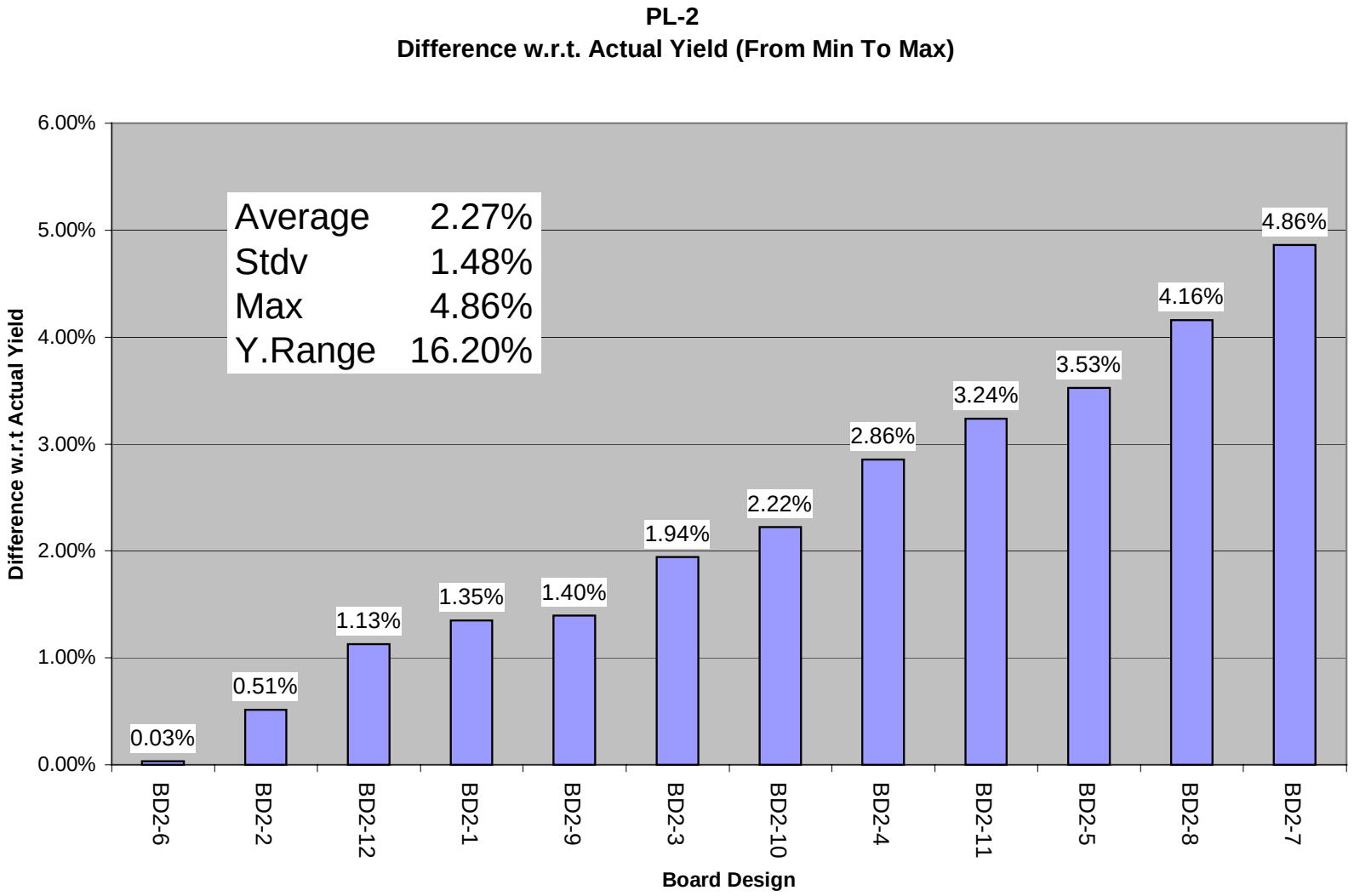

Figure 8.4. Average Difference With Respect to Actual Yield for PL-2

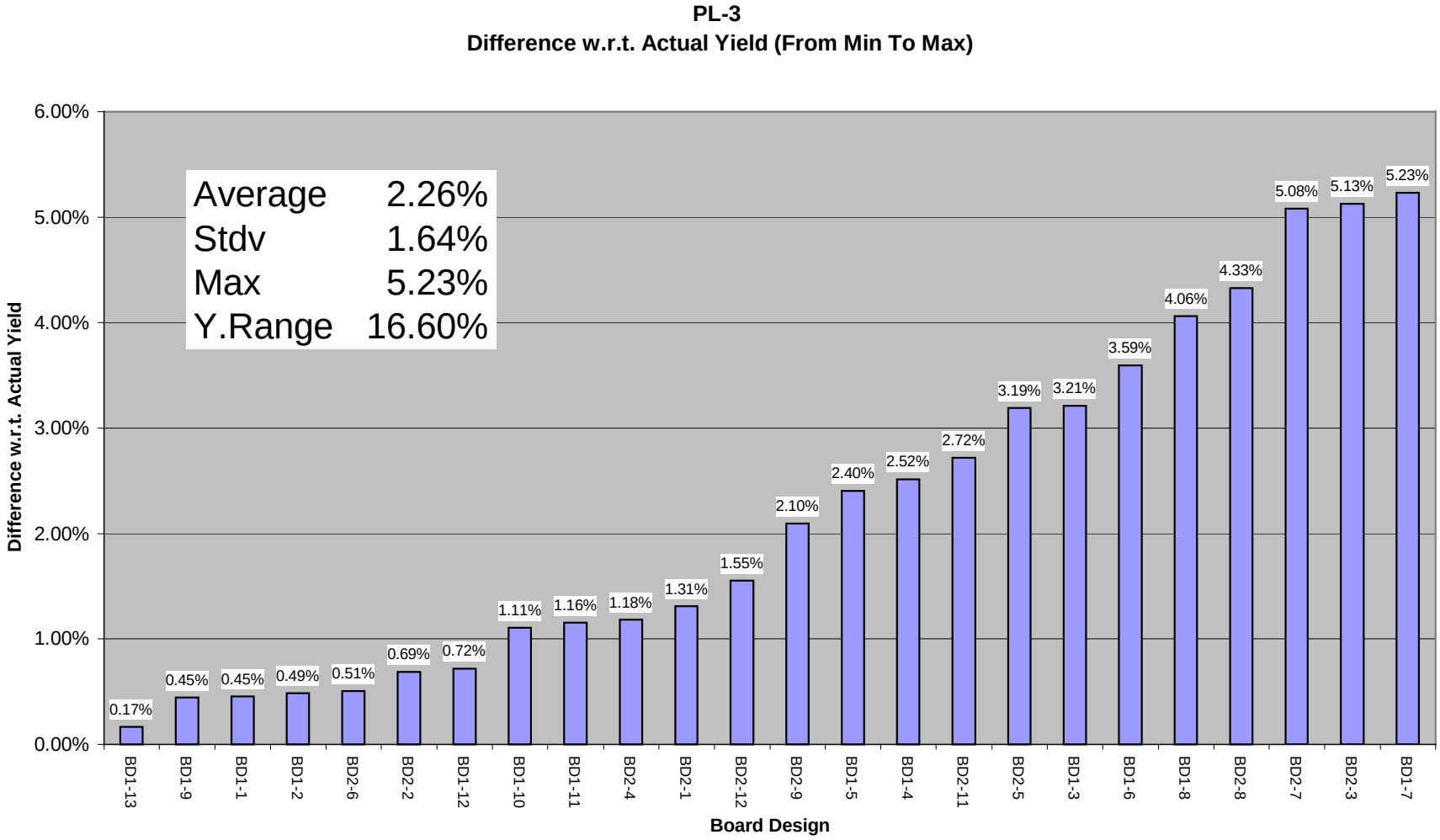

Figure 8.5. Average Difference With Respect to Actual Yield for PL-3

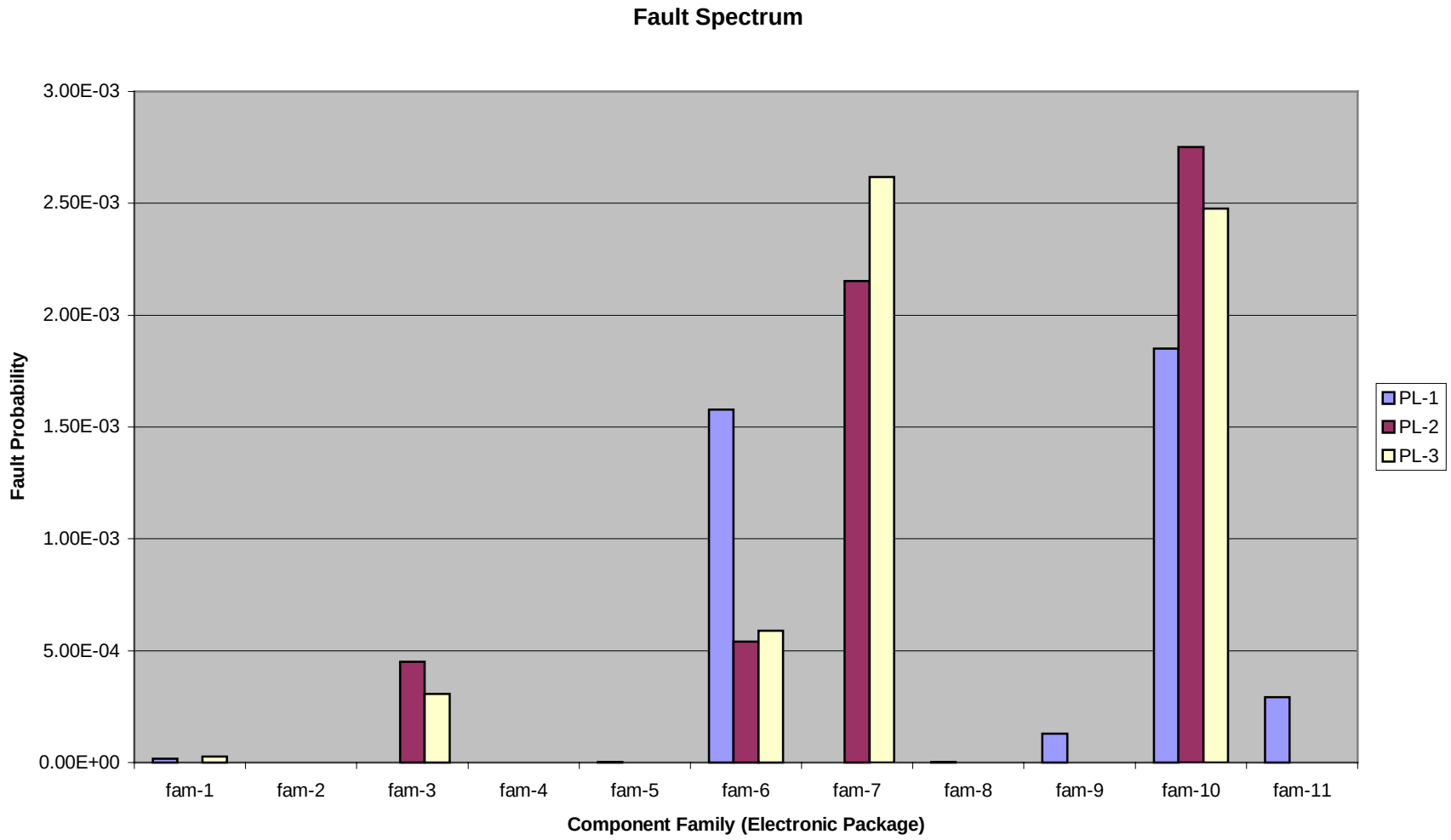

Figure 8.6. Fault Spectrum per Component Families for PL-1, PL-2 and PL-3

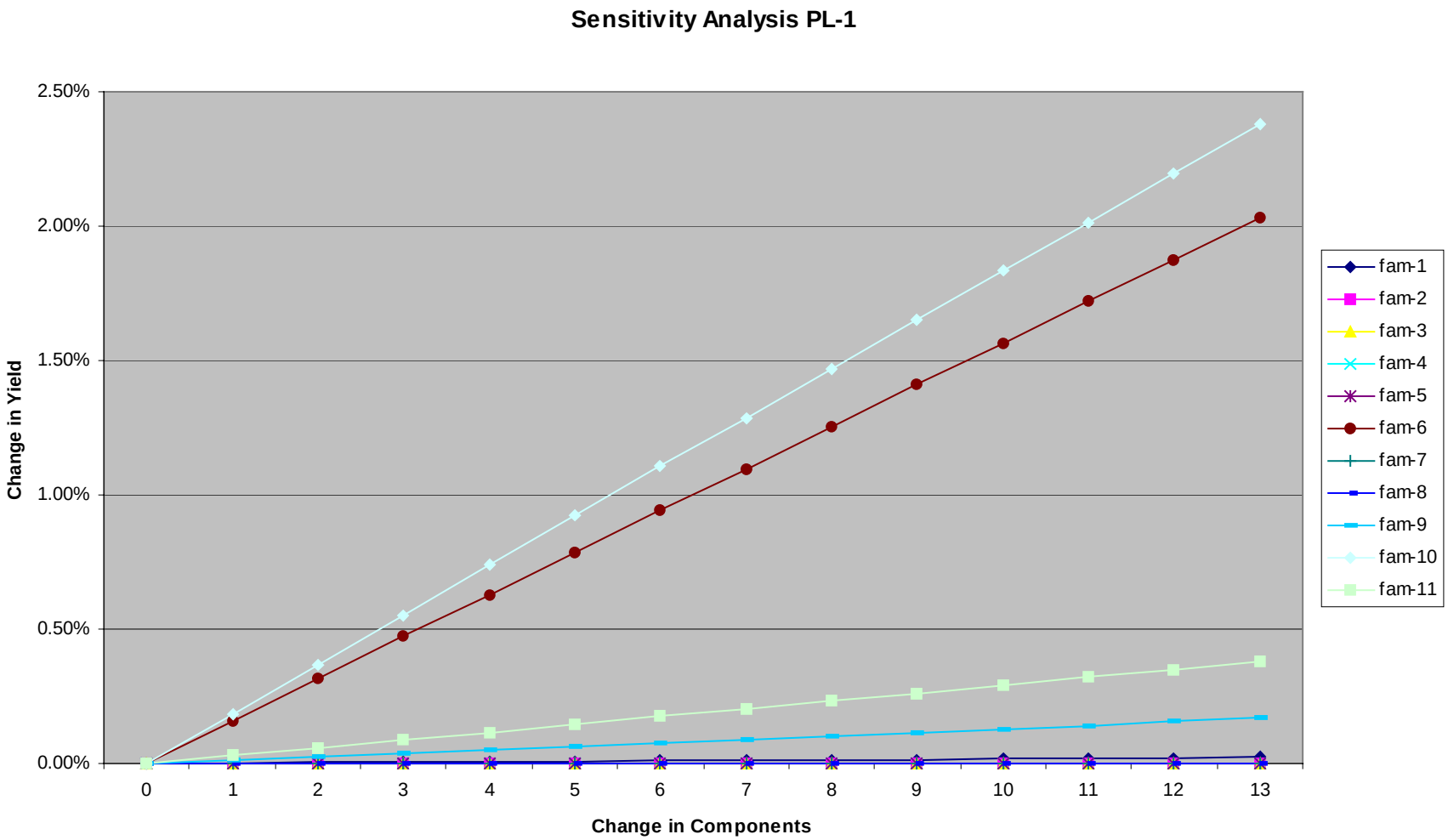

Figure 8.7. Case Study Sensitivity Analysis for PL-1

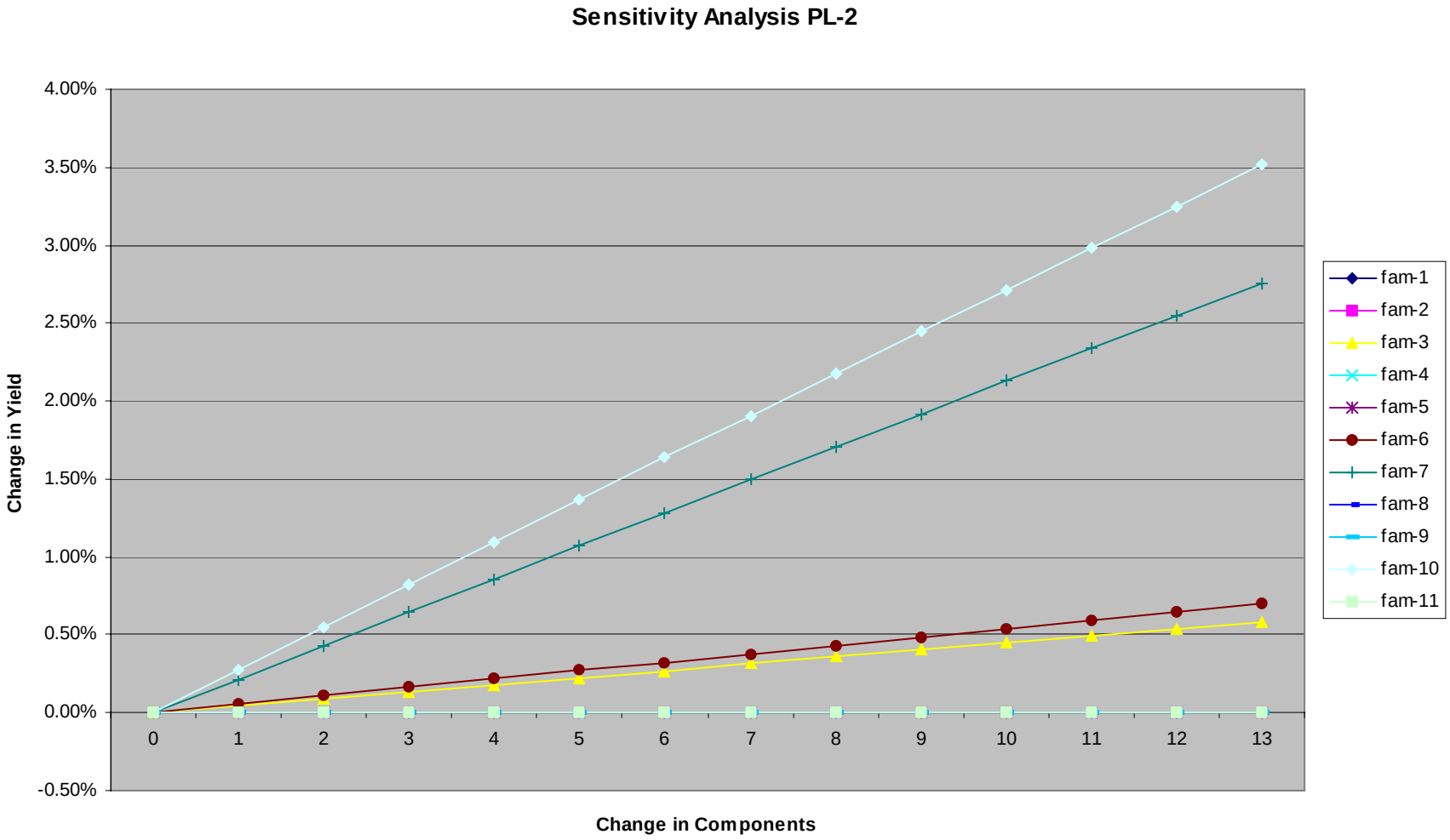

Figure 8.8. Case Study Sensitivity Analysis for PL-2

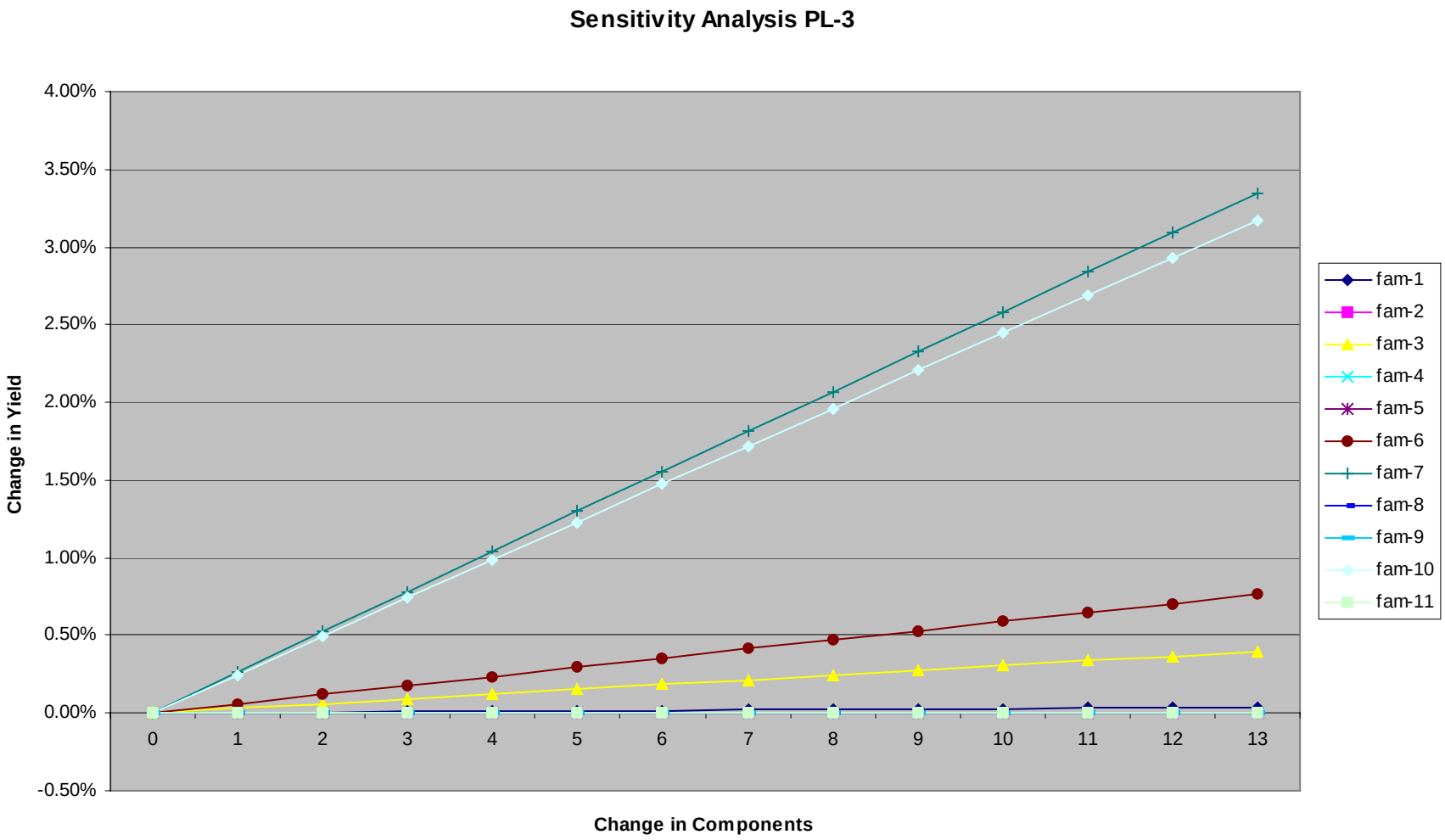

Figure 8.9. Case Study Sensitivity Analysis for PL-3

8.5 Case Study Result Analysis and Conclusions

The results from this case study clearly show the applicability of the yield model to actual production lines. The results presented in Tables 8.1 through 8.3 illustrate how the yield predictions from the model are on average within 1.25% from the actual yield for PL-1, 2.27% for PL-2, and 2.26% for PL-3. The standard deviations of these results were determined to be 1.24% for PL-1, 1.48% for PL-2, and 1.64% for PL-3. In addition, the maximum difference was 4.33% for PL-1, 4.86% for PL-2, and 5.23% for PL-3. These numbers confirm that the yield model presented in this research is a reliable tool that can be used to estimate the manufacturing yield of any given PCB production line given that past performance data is available.

These results can be of particular use for our corporate sponsor since they have now determined the component type fault probability spectrums for two of their production lines. These spectrums can in turn be used to predict the yield of new board designs. Finally, the results from the sensitivity analysis suggest that the production lines in the study have some difficulties assembling board designs that contain a high number of components of fam-7 and fam-10. This information can assist PCB design engineers when considering manufacturability issues for their designs. Multiple "what-if" analysis can be considered (similar to those explained in the previous chapter) to determine the board design best suited for each line.

CHAPTER IX

CONCLUSIONS AND FUTURE RESEARCH

This thesis has presented the development of a new mathematical model for predicting the manufacturing yield of PCB assembly lines. Conclusions have been drawn throughout the development of the model from its original formulation to its final case study implementation. This chapter summarizes some of these conclusions and suggests future research directions in this field.

After performing a thorough literature review of the research done in the area of manufacturing yield, it was determined that there is no tool available for assisting engineers in determining reliable estimates of their production capabilities as they introduce new board designs onto their current production lines. Motivated by this need, a more in-depth study of manufacturing yield as well as the electronic assembly process was undertaken. The relevant literature research was divided into three main fields: process modeling, board design and PCB testing. Process modeling accounts for past yield models developed in the early 1980's for the IC industry. These models are usually probability-based models that require fault probability inputs in order to estimate yield. The Williams and Brown model, simplified to a Poisson model, is the most well-known of these models. Board design yield models account for past PCB production line performance and aim at fitting empirical models (regression lines or artificial neural networks) to estimate design parameters that could then be used to predict yield. Finally, the PCB testing literature portrays the relationship that exists between testing and yield and exemplifies how an increase in one could have a direct effect on the other. The

model presented in this research combines elements from process modeling and board design into a single yield model.

An optimization model was formulated (using the Poison model from process modeling) where fault probabilities were optimized to minimize the difference between actual yield values and predicted yield values (using the past yield history and design information from board design). A preliminary case study was developed to test the model. Data for this case study was available in the literature and accounted for 30 different board designs with actual production line yields for each. The optimization model was solved using a genetic algorithm approach. Three scenarios were considered each accounting for a different number of design parameters (component family types). The first scenario considered all components to be the same; the second divided them according to their assembly process (through-hole or surface-mount), and finally the last scenario considered each component family according to the type of leads. After solving the optimization model for each scenario, it was concluded that:

- The more design characteristics (component family types) considered, the more accurate the yield model.
- The more design characteristics considered, the more challenging and time consuming the optimization model (adding one dimension to the solution space for every additional component family considered).
- The proposed yield model predicts yield within 3% of the actual yield value - outperforming previous regression models (within 10%), and artificial neural network models (within 5%).

Furthermore, the yield model was revised and its underlying assumption regarding fault formation was validated. This was done by considering a case in which fault formation in the boards was not independent of previous faults already on the board. The same case study was performed, but a new optimization model was formulated in order to account for the dependant fault formation. The model used was based on Stapper's negative binomial model developed in the 1980's for the IC industry. From the results of this analysis, it was concluded that:

- Faults do not appear to be dependant of previous faults already on the board. For the three scenarios considered, the original Poisson optimization model outperformed the negative binomial model. Although this is not a conclusive proof that faults are independent, it strongly suggests that they are.

Once the Poisson yield model was shown to provide better yield predictions than previous models, the model was reformulated in order to simplify its applicability. As mentioned before, the original Poisson optimization model was solved using genetic algorithms. Due to the nonlinear nature of this optimization problem these algorithms take a considerable long amount of time to reach a solution. The reformulated model was tested using different search algorithms as well as different software packages. From the results of these tests it was concluded that:

- A search algorithm that executes multiple quasi-Newton and conjugate gradient searches provided the best solution based on the time it took for the search to converge to a solution, and the accuracy of the results. This algorithm was coded and implemented using Microsoft's Visual Basic for Applications for Excel.

Two additional constraints were considered once the model was reformulated. The first constraint accounted for cases in which yield was never to be over estimated. The second case considered the under estimation constraint. The addition of these constraints into the model transforms the optimization problem into a simple linear program. From the results of these two cases, it can be concluded that:

- The addition of these constraints makes the problem considerably easier to implement, noticeably reducing the required solution time.
- The results from the constraint models do not predict yield as accurately as the non-linear model. The average difference with respect to the actual yield increased to 10.4% and 6.6% for the over and under estimation cases, respectively. Therefore, these models should only be considered for specific problems in which the constraints must be accounted for.

In addition to predicting yield, this research also focused on providing production engineers with a tool that could potentially assist them in determining key components that affect production capabilities. A sensitivity analysis was performed on each of the component types considered for the preliminary case studies. Different board designs were evaluated and analyzed to show the applicability of the model and its results in the process of estimating yield. It was concluded from this analysis that:

- Certain component types may have a considerably higher effect on yield than others (such as JD and nsmthd in the preliminary case study).
- Once the yield model is implemented, design suggestions can be made in order to account for manufacturability issues at the design level.

The last step in the development of the presented yield model was to test it using new data from actual production lines. A questionnaire was developed and circulated around the world with the assistance of a corporate sponsor. Based on the feedback from the questionnaire a suitable surface mount assembly facility line was selected for the case study. Data from two current production lines was provided and the reformulated Poisson model was selected to be the most suited for the case study. The multiple quasi-Newton and conjugate gradient search algorithms were used to solve the optimization problem. In addition to the two lines considered, a third line comprised of the aggregate of the two original lines was analyzed as well. From this case study it was determined that:

- The proposed yield model continues to provide very good yield predictions. The results from the actual case study ranged between 1.25% and 2.27% for the three lines in study.
- Through sensitivity analysis of the component families, two families were determined to be the ones most affecting the manufacturing yield of the boards.

In summary, the yield model presented in this research has been developed, validated, refined, reformulated, implemented, tested and analyzed successfully. Future applications that utilize the developed model should build on the concepts presented in this research. Most attractive of all is the fact that the model requires very little data collection (most of which is already available) in order to be implemented. The advantages that can come from the model can be significant especially when used to aid in the design process of future PCBs. Design for manufacturing (DFM) has been a popular area of research for many years. An interesting application of the proposed yield

model is to see how it can become a Design for “specific line” Manufacturing tool. As more data is available, the production lines become more and more tailored, making the yield model estimates more reliable.

This study could also be expanded to consider the cases in which old board designs are produced on new production lines and the cases where new board design are to be produced on new production lines. PCB testing is another vast research area that can be studied in much more depth especially by focusing on its relationship to manufacturing yield (i.e., testing effectiveness and fault coverage).

Finally, another future research development of this yield model would be to combine it with a production cost model. If done appropriately, production engineers could set target costs and determine what kind of production yields are needed to meet these costs. Targeted costs then imply targeted yields, and targeted yields will imply targeted fault probabilities that can be solved for per component type using the yield model presented in this research. In this case actual yields will be accounted for as targeted yields. Once targeted fault probabilities are determined then equipment can be purchased that meets these probability constraints. In essence, this combination yield and cost model will portray the basis of a very useful decision support system for the electronic assembly process of printed circuit boards.

REFERENCES

1. Aas, E.A., *A closer look at multiple faults and defect levels in Digital Circuits*, International Society for Computers and Their Applications ISCAS'88 , 1988, p. 441-444.
2. Aas, E.J. and V.T. Minh, *Defect level calculation: The importance of accurate models for defect Distribution and Multiple Fault coverage in low yield situations*, IEEE- International Society for Computers and Their Applications ISCA'89, 1989, p. 939-944.
3. Agrawal, V.D., S.S. Seth, and P. Agrawal, *Fault coverage requirements in production testing of LSI circuits*, IEEE Journal of Solid-State Circuits, 1982, Vol. SC-17, No. 1, p. 57-61.
4. Alunkal, J.M., *Designing and Integrated Corporate Test Strategy*, Electronics Design, 1993, March, p. 81-89.
5. Ambler, T., R.G.B. Bennetts, H. Bleeker, and G. O'Donnell, *The Economics of Design for Test*, EE- Evaluation Engineering, 1994, December, p. 22-26.
6. Bazaraa, M., H. Sherali, and C. Shetty, *Nonlinear Programming Theory and Algorithms*. 1993, New York: John Wiley & Sons.
7. Bauer, C.E. and J.M. DeCarlo, *Precision Placement of High Lead Count Packages*, ISHM '91 Proceedings, 1991, p. 333-338.
8. Baumgartner, D., *The designer's view: Optimizing manufacturing yields*, Printed Circuit Design, 1997, Jan, p. 19-24.
9. Bloch, C. and R. Ranganathan, *Process Based Cost Modeling*, IEEE/CHMT Symposium, 1991, p. 406-412.

10. Brennan, R.W. and P. Rogers, *Stochastic Optimization Applied to a Manufacturing System Operation Problem*, Proceedings of the 1996 Winter Simulation Conference, Washington, D.C. , 1996, p. 857-864.
11. Chevalier, P.B. and L.M. Wein, *Inspection for Circuit Board Assembly*, Management Science, 1997, Vol. 43, No. 9, p. 1198-1213.
12. Chiluvuri, V.K.R. and I. Koren, *New routing and compaction strategies for yield enhancement*, IEEE-International Workshop on Defect and Fault tolerance in VLSI Systems, 1992, p. 325-335.
13. Clark, A.J., A.J. White, P.G. Leany, and A.J. Wycherley, *A Quality Modeling System for Predicting the Yield of Assembly and Test Processes in the Printed Circuit Assembly Industry*, Journal of Electronics Manufacturing, 1995, Vol. 5, No. 2, p. 75-87.
14. Coombs, C., *Printed Circuits Handbook*. 1988, New York: McGraw-Hill.
15. Corsi, F., S. Martino, and T.W. Williams, *Defect level as a function of fault coverage and yield*, IEEE- Proceedings of the Third European Test Conference, 1993, p. 507-508.
16. deSousa, J.T., F.M. Goncalves, J.P. Teixeira, C. Marzocca, F. Corsi, and T.W. Williams, *Defect level evaluation in an IC design environment*, IEE Transactions on Computer-aided design of Integrated Circuits and Systems, 1996, Vol. 15, No. 10, p. 1286-1293.
17. Dooley, B.J., *A Model for the Prediction of Assembly, Rework, and Test Yields*, IBM Journal of Research and Development, 1983, Vol. 27, No. 1, p. 59-67.

18. Driels, M. and J.S. Klegka, *Analysis of Alternative Rework Strategies for Printed Wiring Assembly Manufacturing Systems*, IEEE Transactions on components, hybrids, and manufacturing technology, 1991, Vol. 14, No.. 3, p. 637-644.

19. Graham, G., *Achieving Perfect SMT Component Location After Reflow*, Surface Mount Technology, 1990, September, p. 51-54.

20. Gujar, R., P. Desruelle, and R. Hammer, *Electronic Manufacturing Cell Design Using PC-Based Modeling and Simulation Tools*, Proceedings of the 1990 Winter Simulation Conference, 1990, p. 559-563.

21. Henshaw, B., *How much fault coverage is enough?*, 1989 International Test Conference, 1989, p. 939-940.

22. Holden, H., *Predicting cost tradeoffs in design for manufacturing (DFM)*, SMI: Proceedings of the Technical Program SMT, 1995, p. 659-665.

23. Hotchkiss, J., *How much testing should be done in-circuit?*, Electronics Industry, 1980, p. 19.

24. Hroundas, G., *Economics of bare printed circuit board testing*, Circuit World, 1986, Vol. 12, No. 2, p. 31-34.

25. Hroundas, G., *New directions in electrical testing of PCBs for the 1990's*, Circuit World, 1990, Vol. 16, No. 4, p. 29-31.

26. Joseph, L.A., J.T. Watt, and N. Wigglesworth, *Modeling and Analysis of a New Product Development in Electronics Sub-Assembly Manufacturing*, Proceedings of the 1990 Winter Simulation Conference, 1990, p. 899-903.

27. Kuhn, K., *Key Issues in Surface Mount Assembly*, Surface Mount Technology, 1990. July, p. 13-15.

28. Lasky, R., *Secondary processes impact throughput, yield and cost*, Electronic Packaging & Production, 1998, p. 119-122.

29. Li, Y., R.L. Mhajan, and J. Tong, *Design Factors and their effect on PCB assembly yield- Statistical and neural network predictive models*, IEEE-CHMT International Electronics Manufacturing Technology Symposium, 1993, p. 353-361.

30. Li, Y., *The effect of printed circuit board design factors on assembly yield*, Ph.D. Dissertation *Department of Mechanical Engineering*. 1993, University of Colorado, p. 100.

31. Li, Y., R.L. Mhajan, and J. Tong, *Design Factors and their effect on PCB assembly yield- Statistical and neural network predictive models*, IEEE Transactions on Components, Packaging and Manufacturing Technology Part A, 1994, Vol. 17, No. 2, p. 183-191.

32. Liljegren, J.D., *Modeling Final Assembly and Test Process in the Semiconductor Industry*, Proceedings of the 1992 Winter Simulation Conference, 1992, p. 856-860.

33. Linn, R.J., *Component placement error model for surface mounted technology*, 5th Industrial Engineering Research Conference Proceedings, 1996, p. 91-96.

34. Matheson, R., *Second generation PCB self learn*, ATE'87, 1987, p. 78-88.

35. Millman, S.D., *Improving Quality: Yield Versus Test Coverage*, Journal of Electronics Testing, 1994, Vol. 5, p. 253-261.

36. Moyer, L.K. and S.M. Gupta, *An efficient assembly sequencing heuristic for Printed Circuit Boards configurations*, Journal of Electronics Manufacturing, 1997, Vol. 7, No. 2, p. 143-160.

37. Nachlas, J.A., S.R. Loney, and B.A. Binney, *Diagnostic-strategy selection for series systems*, IEEE Transactions on Reliability, 1990, Vol. 39, No. 3, p. 273-280.

38. Orsejo, S., *Test Strategies for PCB Manufacturing*, Future Circuits International, 1998, No. 3, p. 235-242.

39. Park, E.S., M.R. Mercer, and T.W. Williams, *A statistical model for delay-fault testing*, IEEE Design & Test of Computers, 1989, Feb., p. 45-55.

40. Park, E.S., M. Ray, and T.W. Williams, *The total delay fault model and statistical delay fault coverage*, IEEE Transaction on Computers, 1992, Vol. 41, No. 6, p. 688-698.

41. Phillips, J.C., *Essential testability guidelines for current technology*, IEEE, 1993, p. 273-282.

42. Pynn, C.T., *Manufacturing process test: A strategy for the TQM environment*, NEPCON-National Electronic Packaging and Production Conference WEST'93, 1993, p. 895-903.

43. Pynn, C.T., *Combine test steps to reduce test costs*, Test & Measuremnt World, 1995, May, p. 39-46.

44. Pynn, C.T., *Vectrorless test boosts fault coverage and cuts cycle time*, EE-Evaluation Engineering, 1995, August, p. 37-41.

45. Pynn, C.T., *Shriniking the Cost of Board Test*, EE-Evaluation Engineering, 1996, June, p. 38-45.

46. Pynn, C.T., *Analyzing Manufacturing Test Costs*, IEEE Design and Test of Computers, 1997, July, p. 36-40.

47. Rowland, R. and P. Belangia, *Applied Surface Mount Assembly*. 1993, New York: Van Nostrand Reinhold.

48. Shih, W., Y. Huang, and K. Srihari, *Yield Estimation Through Component Placement Accuracy Evaluation in Surface Mount PWB Assembly*, Journal of Electronics Manufacturing, 1996, Vol. 6, No. 1, p. 51-63.

49. Sousa, J.T., T. Shen, and P.Y.K. Cheung, *Realistic Fault Extraction for Boards*, IEEE Computer Society Press, 1996, p. 612.

50. Stapper, C.H., *Integrated Circuit Yield Statistics*, Proceedings of the IEEE, 1983, Vol. 71, No. 4, p. 453-470.

51. Stapper, C.H., *Simulation of spatial fault distributions for integrated circuit yield estimations*, IEEE Transaction on Computer –Aided Design, 1989, Vol. 8, No. 12, p. 1314-1318.

52. Stapper, C.H., *Fault Simulation programs for integrated circuit yield estimations*, IBM Journal of Research and Development, 1989, Vol. 33, No. 6, p. 647-652.

53. Stern, O. and H. Wunderlich, *Simulation results of an efficient defect analysis procedure*, IEEE- International Test Conference 1994, 1994, p. 729-738.

54. Tegethoff, M.V., T.E. Figal, and S.W. Hird, *Board Test DFT Model for Computer Products*, International Test Conference 1992, 1992, p.367-371.

55. Tuttle, J.A., T.W. Collins, and M.S. Tuttle, *Matching models to real life for defect reduction*, IEEE-International Test Conference, 1995, p. 217-223.

56. Wadsack, R.L., *Fault coverage in digital integrated circuits*, Bell System Technological Journal, 1978, Vol. 57, No. May-June, p. 1475-1488.

57. Wadsack, R.L., *VLSI: How much fault coverage is enough?*, Proceedings of the International Test Conference, 1981, p. 547-554 .

58. Wang, M.J., J.E. Chen, and Y.Y. Chen, *To verify manufacturing yield by testing*, IEEE Computer Society Press, 1994, p. 385-390.

59. Williams, T.W. and N.C. Brown, *Defect Level as a Function of Fault Coverage*, IEEE Transactions on Computers, 1981, Vol. C-30, No. 12, p. 987-988.

60. Williams, T.W., *Test length in a self-testing environment*, IEEE Design &Test, 1985 April, p. 59-63.

APPENDIX A

COMPONENT TYPES

This Appendix outlines the most common electronic components used in today's electronic assembly industry. According to Li [Li, 1994] five major factors affect manufacturing yield (presented in Chapter 2-2). Two of this factors, (1) package types of the components on the board, and (2) different types of leads, will be classified in more detail. This outline is taken from Li [Li, 1994]:

A.1. Package Types of the Components on the Board

Manufacturing yield strongly depends on component placement and insertion [Shih, 1996], therefore understanding that there are differences among component packages is important when modeling yield.

Through-hole components (see Figure A.1):

a) Axial and radial leaded through hole components

b) Dual-in-line and single-in-line packages

c) Pin Grad Array (PGA)

d) Non-standard through hole devices (sockets, connectors, etc.)

Surface Mount components (see Figure A.1):

e) Chip capacitors and resistors

f) Transistors (SOTs)

g) Plastic Leaded Chip Carriers (PLCCs)

h) Small Outline Integrated Circuits (SOICs)

i) Lead-less Chip Carriers (LCCs)

j) Quad Flat Pack (QFP)

Fine Pitch packages with pitch 25 mil. or less. Main groups,

k) Quad Flat Pack (QFP)

l) The guard-ring packages

m) The small outline (SOP) packages

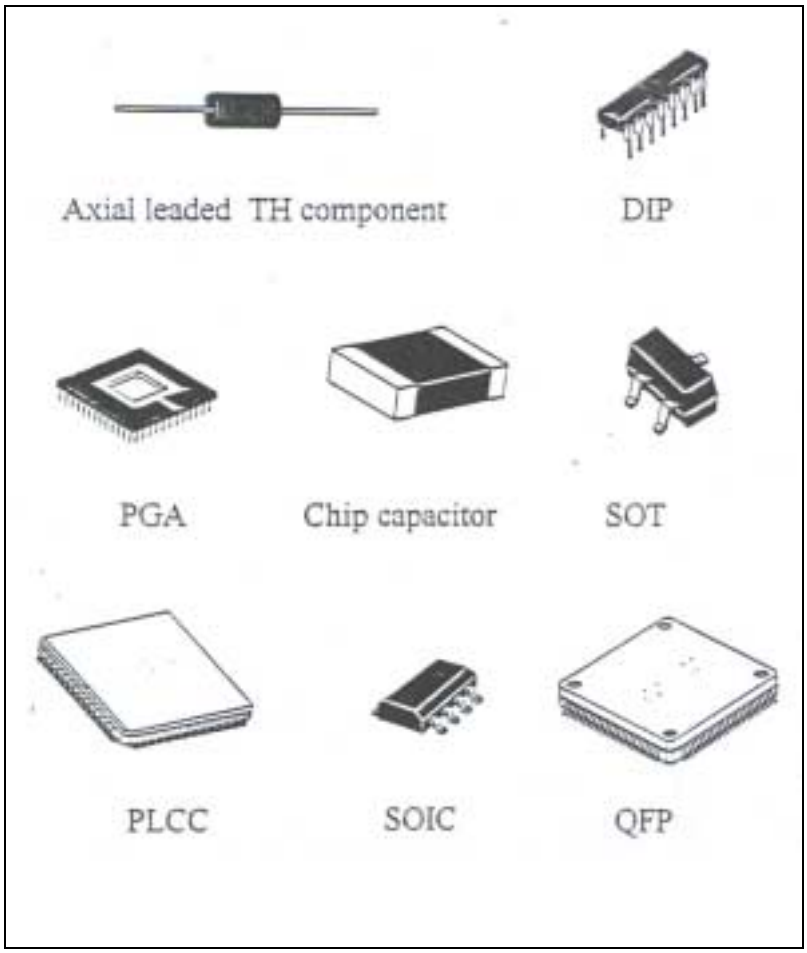

Figure A.1 Illustration of some electronic packages [Li, 1994]

A1. 2. Different Types of Leads

Differentiating among the different types of leads is important because they affect the PCB's solder printing and soldering characteristics.

Through Hole leads:

a) Axial and radial leads

b) Leads of DIPs and SIPs

c) Leads of PGAs

d) Non standard through hole leads

Surface Mount Leads:

e) Termination of chip capacitor and resistor

f) Butt leads

g) J-Leads

h) Gull-wing leads of SOICs

i) Leads of QFPs

Fine-pitch leads:

j) Leads of QFPs

k) Leads of guard-ring packages

l) Leads of SOPs

APPENDIX B

RESULTS FROM PRELIMINARY CASE STUDY

Model I= Poisson model

Model II = Negative-Binomial Model

Tables Included:

Table B.1 Results for Single-Variable Case Using Model I

Brd #	N	Yactual	Y model	Difference
1	3517	86.32%	80.13%	7.18%
2	2303	80.64%	86.49%	7.26%
3	2214	84.10%	86.98%	3.43%
4	1451	88.23%	91.26%	3.44%
5	1177	88.44%	92.85%	4.99%
6	1954	93.09%	88.42%	5.02%
7	2203	96.05%	87.04%	9.38%
8	1655	90.30%	90.10%	0.22%
9	512	98.91%	96.83%	2.11%
10	1782	96.65%	89.38%	7.52%
11	682	99.91%	95.79%	4.12%
12	548	97.61%	96.61%	1.03%
13	1602	98.76%	90.40%	8.47%
14	1481	95.70%	91.09%	4.82%
15	532	99.37%	96.70%	2.68%
16	762	98.14%	95.31%	2.88%
17	824	97.20%	94.94%	2.32%
18	1705	91.29%	89.81%	1.62%
19	1391	92.82%	91.61%	1.30%
20	4302	79.49%	76.26%	4.06%
21	4312	61.74%	76.21%	23.44%
22	3040	82.57%	82.57%	0.00%
23	4994	61.67%	73.01%	18.38%
24	449	88.48%	97.21%	9.87%
25	816	97.61%	94.99%	2.68%
26	1078	90.61%	93.43%	3.12%
27	2291	81.43%	86.56%	6.30%
28	1773	88.84%	89.43%	0.67%
29	1336	97.86%	91.93%	6.06%
30	3536	68.14%	80.03%	17.45%

Table B.2 Results for Two-Variable Case Using Model I

Brd #	SMT	TH	Yactual	Y model	Difference
1	2612	905	86.32%	75.68%	12.33%
2	1452	851	80.64%	84.89%	5.27%
3	1662	552	84.10%	83.80%	0.36%
4	1192	259	88.23%	88.40%	0.20%
5	978	199	88.44%	90.41%	2.23%
6	1523	431	93.09%	85.21%	8.46%
7	1820	383	96.05%	82.87%	13.72%
8	555	1100	90.30%	92.09%	1.99%
9	68	444	98.91%	98.22%	0.70%
10	68	1714	96.65%	95.09%	1.61%
11	0	682	99.91%	98.28%	1.64%
12	0	548	97.61%	98.61%	1.03%
13	68	1534	98.76%	95.53%	3.27%
14	68	1413	95.70%	95.82%	0.13%
15	0	532	99.37%	98.65%	0.72%
16	0	762	98.14%	98.08%	0.07%
17	0	824	97.20%	97.92%	0.74%
18	555	1150	91.29%	91.98%	0.75%
19	1391	0	92.82%	87.27%	5.98%
20	4302	0	79.49%	65.64%	17.43%
21	4312	0	61.74%	65.57%	6.21%
22	2039	1001	82.57%	79.85%	3.30%
23	4920	74	61.67%	61.67%	0.00%
24	407	42	88.48%	95.99%	8.49%
25	0	816	97.61%	97.94%	0.34%
26	630	448	90.61%	92.95%	2.59%
27	1754	537	81.43%	83.08%	2.03%
28	849	924	88.84%	89.88%	1.18%
29	272	1064	97.86%	94.77%	3.16%
30	3118	418	68.14%	72.92%	7.02%

Table B.3 Results for Multi-Variable Case Using Model I

Brd	TH Leads						SMT Leads				F.Pitch					Ymodel	Difference
#	A	DIP	DIPD	Nsth	nshthd	nsmthd	J	G	Ct	Cb	SOTt	SOTb	GD	JD	Y Actual		
1	123	580	34	202	21	2	136	1024	50	1228	96	78	59	2	86.32%	86.31%	0.01%
2	323	240	18	288	8	32	204	314	60	778	0	96	21	3	80.64%	83.95%	4.10%
3	20	292	18	240	24	0	248	560	746	0	108	0	32	14	84.10%	82.89%	1.44%
4	95	28	1	136	13	8	208	512	442	0	30	0	32	4	88.23%	88.23%	0.00%
5	67	28	1	104	4	12	208	356	390	0	24	0	22	4	88.44%	88.91%	0.53%
6	79	0	0	352	36	0	168	846	0	410	0	99	44	3	93.09%	89.30%	4.07%
7	73	64	1	246	11	0	237	1098	4	472	9	0	61	7	96.05%	86.77%	9.67%
8	134	610	30	356	13	21	68	46	354	0	87	0	3	1	90.30%	90.16%	0.15%
9	196	80	5	168	21	0	68	0	0	0	0	0	0	1	98.91%	98.91%	0.00%
10	696	672	35	346	24	2	68	0	0	0	0	0	0	1	96.65%	96.65%	0.00%
11	0	0	0	682	7	0	0	0	0	0	0	0	0	0	99.91%	99.91%	0.00%
12	278	160	6	110	10	8	0	0	0	0	0	0	0	0	97.61%	97.13%	0.49%
13	460	714	37	360	31	1	68	0	0	0	0	0	0	1	98.76%	97.30%	1.48%
14	436	646	34	331	11	0	68	0	0	0	0	0	0	1	95.70%	97.70%	2.09%
15	142	270	13	120	3	0	0	0	0	0	0	0	0	0	99.37%	99.37%	0.00%
16	336	242	13	184	11	0	0	0	0	0	0	0	0	0	98.14%	99.01%	0.88%
17	118	626	25	80	2	0	0	0	0	0	0	0	0	0	97.20%	99.06%	1.91%
18	184	610	30	356	13	21	68	46	354	0	87	0	3	1	91.29%	90.08%	1.33%
19	0	0	0	0	0	0	648	268	468	0	7	0	34	15	92.82%	85.56%	7.82%
20	0	0	0	0	0	0	2000	1678	20	572	32	0	98	30	79.49%	72.01%	9.41%
21	0	0	0	0	0	0	1724	1904	40	544	19	81	113	50	61.74%	61.74%	0.00%
22	469	464	19	68	1	0	136	1888	0	0	15	0	108	2	82.57%	82.57%	0.00%
23	12	56	2	6	1	0	1564	2578	12	766	0	0	148	43	61.67%	61.67%	0.00%
24	2	0	0	40	1	0	80	184	6	134	0	3	10	3	88.48%	96.53%	9.09%
25	434	154	11	228	13	8	0	0	0	0	0	0	0	0	97.61%	96.74%	0.89%
26	74	206	13	168	11	0	0	184	422	0	24	0	10	0	90.61%	95.10%	4.95%
27	65	0	0	472	39	0	460	668	8	516	102	0	39	12	81.43%	87.35%	7.26%
28	132	612	32	180	11	2	136	172	508	0	33	0	10	2	88.84%	92.40%	4.01%
29	446	554	31	64	8	0	272	0	0	0	0	0	0	4	97.86%	96.12%	1.78%
30	60	280	13	78	3	3	504	1638	304	652	20	0	96	12	68.14%	77.53%	13.78%

Table B.4 Results for Single-Variable Case Using Model II

Brd #	N	Yactual	Y model	Difference
1	3517	86.32%	89.72%	3.93%
2	2303	80.64%	90.04%	11.65%
3	2214	84.10%	90.07%	7.09%
4	1451	88.23%	90.38%	2.44%
5	1177	88.44%	90.54%	2.38%
6	1954	93.09%	90.16%	3.15%
7	2203	96.05%	90.07%	6.23%
8	1655	90.30%	90.28%	0.02%
9	512	98.91%	91.18%	7.82%
10	1782	96.65%	90.23%	6.64%
11	682	99.91%	90.96%	8.96%
12	548	97.61%	91.13%	6.64%
13	1602	98.76%	90.31%	8.56%
14	1481	95.70%	90.37%	5.57%
15	532	99.37%	91.15%	8.27%
16	762	98.14%	90.87%	7.40%
17	824	97.20%	90.81%	6.57%
18	1705	91.29%	90.26%	1.13%
19	1391	92.82%	90.42%	2.59%
20	4302	79.49%	89.56%	12.67%
21	4312	61.74%	89.56%	45.07%
22	3040	82.57%	89.83%	8.79%
23	4994	61.67%	89.45%	45.05%
24	449	88.48%	91.28%	3.16%
25	816	97.61%	90.82%	6.95%
26	1078	90.61%	90.61%	0.00%
27	2291	81.43%	90.04%	10.57%
28	1773	88.84%	90.23%	1.57%
29	1336	97.86%	90.45%	7.57%
30	3536	68.14%	89.71%	31.66%

Table B.5 Results for Two-Variable Case Using Model II

Brd #	SMT	TH	Yactual	Y model	Difference
1	2612	905	86.32%	86.58%	0.30%
2	1452	851	80.64%	87.17%	8.10%
3	1662	552	84.10%	87.03%	3.49%
4	1192	259	88.23%	87.37%	0.97%
5	978	199	88.44%	87.58%	0.98%
6	1523	431	93.09%	87.12%	6.41%
7	1820	383	96.05%	86.94%	9.48%
8	555	1100	90.30%	88.16%	2.37%
9	68	444	98.91%	90.35%	8.65%
10	68	1714	96.65%	90.35%	6.52%
11	0	682	99.91%	100.00%	0.09%
12	0	548	97.61%	100.00%	2.45%
13	68	1534	98.76%	90.35%	8.51%
14	68	1413	95.70%	90.35%	5.59%
15	0	532	99.37%	100.00%	0.63%
16	0	762	98.14%	100.00%	1.90%
17	0	824	97.20%	100.00%	2.88%
18	555	1150	91.29%	88.16%	3.43%
19	1391	0	92.82%	87.22%	6.04%
20	4302	0	79.49%	86.07%	8.28%
21	4312	0	61.74%	86.07%	39.41%
22	2039	1001	82.57%	86.83%	5.16%
23	4920	74	61.67%	85.94%	39.35%
24	407	42	88.48%	88.48%	0.00%
25	0	816	97.61%	100.00%	2.45%
26	630	448	90.61%	88.03%	2.85%
27	1754	537	81.43%	86.98%	6.82%
28	849	924	88.84%	87.72%	1.26%
29	272	1064	97.86%	88.90%	9.16%
30	3118	418	68.14%	86.40%	26.79%

Table B.6 Results for Multi--Variable Case Using Model II

Brd	TH Leads						SMT Leads				F.Pitch					Ymodel	Difference
#	A	DIP	DIPD	Nsth	nshthd	nsmthd	J	G	Ct	Cb	SOTt	SOTb	GD	JD	Y Actual		
1	123	580	34	202	21	2	136	1024	50	1228	96	78	59	2	86.32%	78.72%	8.80%
2	323	240	18	288	8	32	204	314	60	778	0	96	21	3	80.64%	80.71%	0.09%
3	20	292	18	240	24	0	248	560	746	0	108	0	32	14	84.10%	88.20%	4.88%
4	95	28	1	136	13	8	208	512	442	0	30	0	32	4	88.23%	88.23%	0.00%
5	67	28	1	104	4	12	208	356	390	0	24	0	22	4	88.44%	88.60%	0.18%
6	79	0	0	352	36	0	168	846	0	410	0	99	44	3	93.09%	81.95%	11.97%
7	73	64	1	246	11	0	237	1098	4	472	9	0	61	7	96.05%	80.07%	16.64%
8	134	610	30	356	13	21	68	46	354	0	87	0	3	1	90.30%	90.10%	0.22%
9	196	80	5	168	21	0	68	0	0	0	0	0	0	1	98.91%	97.62%	1.31%
10	696	672	35	346	24	2	68	0	0	0	0	0	0	1	96.65%	96.94%	0.30%
11	0	0	0	682	7	0	0	0	0	0	0	0	0	0	99.91%	99.66%	0.25%
12	278	160	6	110	10	8	0	0	0	0	0	0	0	0	97.61%	98.00%	0.40%
13	460	714	37	360	31	1	68	0	0	0	0	0	0	1	98.76%	96.97%	1.81%
14	436	646	34	331	11	0	68	0	0	0	0	0	0	1	95.70%	97.45%	1.82%
15	142	270	13	120	3	0	0	0	0	0	0	0	0	0	99.37%	98.54%	0.84%
16	336	242	13	184	11	0	0	0	0	0	0	0	0	0	98.14%	98.49%	0.36%
17	118	626	25	80	2	0	0	0	0	0	0	0	0	0	97.20%	98.49%	1.32%
18	184	610	30	356	13	21	68	46	354	0	87	0	3	1	91.29%	90.10%	1.31%
19	0	0	0	0	0	0	648	268	468	0	7	0	34	15	92.82%	89.81%	3.24%
20	0	0	0	0	0	0	2000	1678	20	572	32	0	98	30	79.49%	80.31%	1.03%
21	0	0	0	0	0	0	1724	1904	40	544	19	81	113	50	61.74%	80.21%	29.91%
22	469	464	19	68	1	0	136	1888	0	0	15	0	108	2	82.57%	87.63%	6.13%
23	12	56	2	6	1	0	1564	2578	12	766	0	0	148	43	61.67%	79.58%	29.04%
24	2	0	0	40	1	0	80	184	6	134	0	3	10	3	88.48%	84.12%	4.92%
25	434	154	11	228	13	8	0	0	0	0	0	0	0	0	97.61%	97.95%	0.35%
26	74	206	13	168	11	0	0	184	422	0	24	0	10	0	90.61%	90.65%	0.05%
27	65	0	0	472	39	0	460	668	8	516	102	0	39	12	81.43%	80.74%	0.84%
28	132	612	32	180	11	2	136	172	508	0	33	0	10	2	88.84%	89.13%	0.33%
29	446	554	31	64	8	0	272	0	0	0	0	0	0	4	97.86%	97.26%	0.61%
30	60	280	13	78	3	3	504	1638	304	652	20	0	96	12	68.14%	78.80%	15.65%

Table B.7. Result Parameters for All Cases

Single-Variable

ModelParameters:	p(N)
Model I	6.30E-05
Model II	8.38E-03

Two-Variable

ModelParameters:	p(SMT)	p(TH)
Model I	9.79E-05	2.55E-05
Model II	1.17E-02	1.00E-09

Multi-Variable

	TH Leads					
Model Parameters:	p(A)	p(DIP)	p(DIPD)	p(Nsth)	p(nshthd)	p(nsmthd)
Model I	1.90E-05	4.25E-06	1.78E-04	1.32E-06	3.08E-08	2.75E-03
Model II	6.75E-05	2.71E-04	7.02E-04	1.00E-300	3.40E-04	5.98E-04
	SMT Leads					
	p(J)	p(G)	p(Ct)	p(Cb)		
Model I	0.00E+00	4.85E-05	7.03E-05	8.35E-06		
Model II	1.00E-300	4.82E-04	4.10E-06	8.87E-03		
	F.Pitch					
	p(SOTt)	p(SOTb)	p(GD)	p(JD)		
Model I	4.67E-17	1.95E-04	6.85E-04	5.79E-03		
Model II	1.03E-03	1.00E-300	9.56E-03	1.50E-03		

Table B.8 Result Comparison for the Cases in Study

POISSON	*Model 1-1*	*Model 1-2*	*Model 1-Multi*
Mean	5.73%	3.76%	2.91%
Standard Error	1.01%	0.81%	0.69%
Median	4.09%	2.01%	1.39%
Standard Deviation	5.51%	4.45%	3.75%
Sample Variance	0.30%	0.20%	0.14%
Kurtosis	347.12%	234.34%	117.48%
Skewness	185.16%	166.18%	141.40%
Range	23.44%	17.43%	13.78%
Minimum	0.00%	0.00%	0.00%
Maximum	23.44%	17.43%	13.78%
Sum	171.81%	112.92%	87.17%
Count	30	30	30
Largest(1)	23.44%	17.43%	13.78%
Smallest(1)	0.00%	0.00%	0.00%
Confidence Level(95.0%)	2.06%	1.66%	1.40%

STAPPER	*Model 2-1*	*Model 2-2*	*Model 2-Multi*
Mean	9.34%	7.34%	4.82%
Standard Error	2.06%	1.84%	1.48%
Median	6.80%	4.32%	1.17%
Standard Deviation	11.27%	10.08%	8.08%
Sample Variance	1.27%	1.02%	0.65%
Kurtosis	595.88%	598.28%	444.26%
Skewness	252.48%	252.96%	222.74%
Range	45.07%	39.41%	29.91%
Minimum	0.00%	0.00%	0.00%
Maximum	45.07%	39.41%	29.91%
Sum	280.11%	220.29%	144.63%
Count	30	30	30
Largest(1)	45.07%	39.41%	29.91%
Smallest(1)	0.00%	0.00%	0.00%
Confidence Level(95.0%)	4.21%	3.76%	3.02%

APPENDIX C

NEGATIVE BINOMIAL YIELD MODEL

In a paper presented in 1983[Stapper, 1983], Stapper mathematically showed the validity of using the negative binomial as a proper distribution for modeling the yield of ICs. This appendix presents Stapper's mathematical analysis of the yield model.

C.1 Stapper's Mathematical Analysis

According to Stapper the objective of a yield model is to mathematically describe the number of faults per chip occurring during the manufacturing process. The probability distribution function is designated by *p(k,t),* where there are *k* faults per board at time *t*. The probability distribution function of *p(k,t)* is,

$$G(s;t) = \sum_{k=0}^{\infty} p(k,t)s^k$$

where *s* is a complex variable. This function is related to the distribution function *p(k,t)* as follows:

$$p(k,t) = \frac{1}{k!}\frac{\partial^k G(s;t)}{\partial s^k}\bigg|_{s=0}.$$

Note that *p(k=0,t)* represents the yield of the manufacturing process (Y),

$$Y = p(0,t).$$

In addition, Stapper showed that yield can also be expressed as a function of the probability generating function as,

$$Y = p(0,t) = G(0;t).$$

This implies that the distribution generating function *G(s;t)* is as closely related to the manufacturing yield as the fault distribution *p(k,t).* Rogers [Rogers, 1974], showed that *G(s;t)* must also satisfy the equation:

$$\frac{\partial}{\partial t} G(s;t) = (s-1)\sum_{k=0}^{\infty} p(k,t)\, f\,(k,t) s^k$$

where *f(k,t)* is a probability distribution function that describes how the faults are formed on the chip during the manufacturing process. The quantity *f(k,t) dt* is defined as the probability that, during a given time interval *{t ,t+ dt}*, a fault will occur on a board which already has k faults caused by defects. The distribution generating function *G(s;t)* and the fault distribution *p(k,t)* depend completely on the nature of the function *f(k,t).*

C.2 Stapper's Mathematical Derivation of the Negative-Binomial Model

Stapper's Assumption (pg. 43):

> *The probability that a fault of type f_i occurs is independent of time and increases linearly with respect to the number of faults of type f_i that have already occurred.*

In particular, assume that if the fault types are indicated by a subscript *i,* the probability density function of fault formation is given by

$$d_i(k_i,t) = c_i + b_i k_i$$

where c_i, b_i are nonnegative constants, and k_i is the number of faults of type *i* present on the board. Under these conditions

$$\frac{\partial}{\partial t} G_i(s;t) = (s-1)\sum_{k_i=0}^{\infty} p(k_i,t) d_i(k_i,t) s^{k_i} = (s-1)[c_i G_i(s;t) + b_i s \frac{\partial}{\partial s} G_i(s;t)].$$

The solution for this differential equation is

$$G_i(s;t) = [e^{b_i t} * (e^{b_i t} - 1)s]^{-c_i/b_i} = \{1 + \frac{(1-s)\lambda_i}{\alpha_i}\}^{-\alpha_i}$$

where, $\alpha_i = \frac{c_i}{b_i}$ and $\lambda_i = \{e^{b_i t} - 1\}\frac{c_i}{b_i}$.

Since yield is *G(0;t)* then

$$Y_i = (1 + \frac{\lambda_i}{\alpha_i})^{-\alpha_i}.$$

Stapper showed that *G(s;t)* is the probability generating function of the negative binomial distribution using the analysis presented in Figure C.1.

Let,

$$p = e^{b_i t} - 1 \qquad \text{and} \qquad k = b_i / c_i$$

$$G(s;t) = G(s) = (1 + p - ps)^{-k}$$

$$G(s) = (1 + p - ps)^{-k} = \left(\frac{1}{1+p}\right)^{k}\left[1 - \left(\frac{p}{1+p}\right)s\right]^{-k}$$

$$= \left(\frac{1}{1+p}\right)^{k} \sum_{r=0}^{\infty} \binom{k+r-1}{r} \left[\left(\frac{p}{1+p}\right)s\right]^{r}$$

$$= \sum_{r=0}^{\infty} \binom{k+r-1}{r} \left(\frac{p}{1+p}\right)^{r} \left(\frac{1}{1+p}\right)^{k} * s^{r}$$

$$= \sum_{r=1}^{\infty} \binom{k-1}{r-1} \left(\frac{p}{1+p}\right)^{r} \left(\frac{1}{1+p}\right)^{k-r} * s^{r}$$

$$= \sum_{r=1}^{\infty} P(r) * s^{r}$$

Where P(r) is the negative binomial probability mass function.

Figure C.1 Stapper's Derivation of Negative Binomial p.g.f.

The Yield for all of the *N* processes can then be written as

$$Y = \prod_{i=1}^{m} Y_i = \prod_{i=1}^{m} (1 + \frac{\lambda_i}{\alpha_i})^{-\alpha_i} \qquad .$$

VITA:

Felipe Helo

Blacksburg, VA 24060
fhelo@vt.edu

INTERESTS: Decision Support Systems. Discrete Event Simulation. E-Commerce. Genetic Algorithms. Scheduling. Statistical Quality Control. Manufacturing. Mathematical Modeling. GUI Development. Facility Layout. Monte Carlo Analysis. Supply Chains. Automation/Robotics. Ergonomics.

EDUCATION:

1996 - 2000 Virginia Tech Blacksburg, VA
Master of Science in Operations Research (ISE)
Bachelor of Science in Industrial and Systems Engineering

1994 - 1996 King College Bristol, TN
Bachelor of Arts in Applied Science and Mathematics
Minor: **Physics**

Gave presentation at the INFORMS (Philadelphia, PA - November 1999) conference.

Chairman 1997-98, Society of Manufacturing Engineers (SME) Virginia Tech student chapter. Gave presentation on the chapter at the regional SME Conference in Raleigh, NC (October 1997)

Member of the Institute of Industrial Engineers (IIE) since 1997