

1-1-2011

# A Die-level Adaptive Test Scheme for Real-time Test Reordering and Elimination

Kapil Ramesh Gotkhindikar  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/open\\_access\\_etds](https://pdxscholar.library.pdx.edu/open_access_etds)

Let us know how access to this document benefits you.

---

## Recommended Citation

Gotkhindikar, Kapil Ramesh, "A Die-level Adaptive Test Scheme for Real-time Test Reordering and Elimination" (2011). *Dissertations and Theses*. Paper 243.  
<https://doi.org/10.15760/etd.243>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

A Die-level Adaptive Test Scheme for Real-time Test Reordering and Elimination

by

Kapil Ramesh Gotkhindikar

A thesis submitted in partial fulfillment of the  
requirements for the degree of

Master of Science

in

Electrical and Computer Engineering

Thesis Committee:

W. Robert Daasch, Chair

C. Glenn Shirley

Mark Faust

Portland State University

©2012

## **Abstract**

Semiconductor manufacturing companies aim to achieve shortest test times for products while maintaining the product quality. Achieving shortest test times for devices requires multiple updates to the test flow and test content. Test cost varies in direct proportion to production test time required to test chips and detect fails. This thesis presents a method to achieve shortest test times by determining when the updates are needed and what are the changes to the test flow and test content.

This thesis introduces a new Adaptive Test Scheme (ATS). ATS estimates individual test fail rates dynamically, per die, and makes real-time modifications to test order and test contents. ATS computes data-driven test fail rate estimates and uses the estimates to identify the required changes and trigger updates to test flow and test content. ATS uses Bayesian statistics to model the per test fail rates and update the test orders. ATS achieves test time reductions by employing per wafer elimination. ATS also incorporates a simple quality monitor, by resetting the test content at the start of next wafer.

This thesis evaluates the performance of ATS with synthetic data generated by a Monte Carlo method and with production wafer sort data for two manufactured products. The product data results show ATS reduced by 20% the total test-time for one product and by 40% for a second product, with changes in product quality level below industry targets.

## **Acknowledgements**

I would like to take this opportunity to express my gratitude to many wonderful people who have encouraged me during my graduate study at Portland State University. This research would not have been successful without the guidance, help and support of my research advisor Dr. Robert Daasch. His extensive knowledge and research experience in the field of IC test and statistics, helped me not only to achieve the goal but also improve my critical thinking and problem solving skills. I am very thankful to our industrial liaisons at Texas Instruments (TI), Kenneth Butler, John Carulli and Amit Nahar for their valuable industrial insights and the opportunity to do field work as summer intern at TI.

Also, I would like to express my gratitude to Professor Glenn Shirley and Professor Mark Faust for serving on my M.S. Defense Examination Committee. I would like to thank Professor Shirley for all the help and guidance he provided in the two years of my association with the Integrated Circuit Design and Test (ICDT) laboratory. I am very grateful to Semiconductor Research Corporation (SRC) for providing me this research opportunity and necessary funds. Special thanks to my dear friends and ICDT lab colleagues Saurabh, Vivek, Chaitrali, Prachi and Satoshi for all the support and encouragement.

Finally, thanks to my parents for their constant support and blessings without which I would not have been here. They raised me up and gave the courage to face the challenges and the blessings to be successful in each endeavour.

## Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>List of Figures</b>	<b>v</b>
<b>1 Introduction: Need, Contribution and Organization</b>	<b>1</b>
1.1 Need and Emergence of Adaptive IC Test . . . . .	1
1.2 Brief Description and Contribution of this Research . . . . .	4
1.3 Organization of Thesis . . . . .	5
<b>2 Background: Previous work, New Classification and Motivation</b>	<b>7</b>
2.1 Previous work on adaptive test . . . . .	8
2.2 New classification of Adaptive Test techniques . . . . .	10
2.3 Motivation . . . . .	14
<b>3 New Method- Adaptive Test Scheme</b>	<b>17</b>
3.1 Using Bayesian Statistics . . . . .	17
3.2 Adaptive Test Scheme . . . . .	18
3.2.1 Applying Bayesian Statistics for Learning Test Fail Rates . .	18
3.2.2 Test and Test Pattern Elimination . . . . .	26
3.3 Adaptive Test Scheme - Monte Carlo Experiments . . . . .	27
<b>4 ATS Performance Evaluation: Monte Carlo and Production data</b>	<b>30</b>
4.1 Monte Carlo Evaluation of ATS . . . . .	30

4.1.1	ATS Test reordering for synthetic test data . . . . .	30
4.1.2	Effect of choice of seed variance on ATS . . . . .	33
4.1.3	Effect of choice of test elimination trigger on ATS . . . . .	36
4.2	Application to Production data . . . . .	40
4.2.1	Description of Production Data . . . . .	40
4.2.2	Reordering Tests and Patterns . . . . .	42
4.2.3	Eliminating Tests and Patterns . . . . .	46
<b>5</b>	<b>Conclusions</b>	<b>51</b>
5.1	Contributions of this work . . . . .	51
5.2	ATS Production Requirements . . . . .	52
5.3	Recommendations and Implications . . . . .	53
	<b>References</b>	<b>55</b>

## List of Figures

1.1	Cost of test and cost of manufacturing . . . . .	1
2.1	Classifying ATS in ITRS format . . . . .	12
2.2	ITRS standard color code. . . . .	12
2.3	ATS test time variance reduction . . . . .	16
3.1	Two-stage Bayesian method . . . . .	20
3.2	Flowchart for Adaptive Test Scheme . . . . .	22
3.3	ATS test fail rate update example . . . . .	25
4.1	Synthetic wafer test fail rates . . . . .	30
4.2	Monte Carlo ATS test reorder example colormap . . . . .	31
4.3	Synthetic wafer test fail rates seed variance example . . . . .	34
4.4	ATS fail rate distributions narrow variance seed . . . . .	35
4.5	ATS fail rate distributions wide variance seed . . . . .	36
4.6	Effect of shock on truncation trigger example . . . . .	39
4.7	ATS reorder only results for Product 1 . . . . .	43
4.8	ATS reorder only results for Product 2 . . . . .	43
4.9	Correlation between ATS test orders and a characterization order . . . . .	44
4.10	Effect of yield on ATS . . . . .	46
4.11	Correlation between consecutive test orders on a wafer . . . . .	47
4.12	ATS Test time vs. Defect level tradeoff Product 1 . . . . .	49
4.13	ATS Test time vs. Defect level tradeoff Product 2 . . . . .	49

# Chapter 1

## Introduction: Need, Contribution and Organization

### 1.1 Need and Emergence of Adaptive IC Test

Semiconductor manufacturing companies aim to achieve a very cost effective manufacturing environment to maximize their profits. Continual device scaling and integration reduces chip manufacturing cost but it does not diminish cost of testing these chips. Aggressive scaling introduces process variation in addition to the equipment variation making it harder to test the same circuits [1,2]. Test cost does not directly scale with transistor count, die size, device pin count, or process technology. Figure 1.1 shows the trends in test cost and silicon manufacturing cost of semiconductor devices. For products in some market segments, test may account for more than 70% of the total manufacturing cost [3].

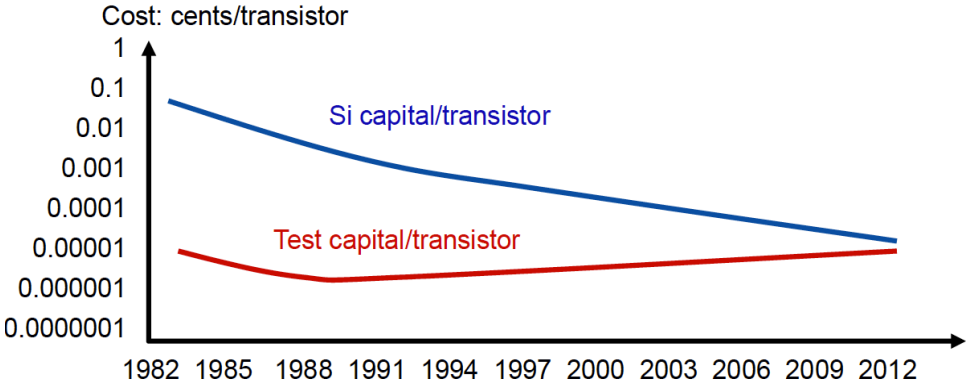


Figure 1.1: Cost of test vs. cost of manufacturing semiconductors from ITRS 2001 [1]



The chip manufacturers are constantly attempting to improve product quality by minimizing number of faulty chips that escape their testing procedures. These faulty chips are called test escapes. Chip manufacturers have to make reliable products and guarantee a certain quality level by limiting the number of test escapes. This quality level is measured in terms of defective parts per million (DPPM) shipped to the customer. Today these quality requirements have become increasingly stringent with target DPPMs being continually driven downward [3]. This results in cost of test being a significant component in the manufacturing cost of the chip. Test cost reduction has become a crucial issue in the overall cost control of the product.

Test cost varies in direct proportion to production test time. Production test time for a failing part is the time required to detect the first fail. This depends on the test application order or test flow and the location of the test that fails the die in the test flow. Achieving shortest test times for devices requires multiple updates to the test flow and test content. This thesis presents a method to achieve shortest test times by determining when the updates are needed and what are the changes to the test flow and test content.

Equipment and process variation leads to various test inadequacies such as high test data volumes, longer test application times, increasing number of test patterns and repetition of tests at different test insertions (like wafer-sort, package, burn-in and final). These test inefficiencies lead to increased test time and test cost. Process and equipment variations result in a non-uniform defect occurrence probability. Moreover, variations result in a highly skewed defect detection probability of the tests or varying test fail rates [4]. The variation in test fail rates suggests emergence

of an approach that can enable test cost reduction by applying a reduced set of test while quantifying associated quality tradeoff [5,6].

Semiconductor manufacturing variation is either spatial or temporal variation. Variation appears on three different scales: lot-to-lot variation, wafer-to-wafer variation and die-to-die variation. Spatial variations are caused by non-uniform process while temporal variations are caused by the drift in process over a period of time [1]. Traditionally, temporal and spatial variation were managed by using proper test and design margins, while defects were screened at test. Today, both variations and defects must be screened at test [2]. When a product is introduced, the initial test content and test order is based on experience, previous similar products and fault simulation. As test fallout data accumulates periodic test program releases are employed to reorder and eliminate tests and test content. Besides being labor-intensive the above method is unresponsive to short term temporal variation or spatial variation.

Over the last few years, adaptive test has been proposed as a test strategy to simultaneously reduce the effect of product variation and screen defects. Adaptive test is recognized as a key driver for future of semiconductor test and was formally defined in International Technology Roadmap of Semiconductors (ITRS), 2009, Test and Test Equipment section [3]. “Adaptive Test is a broad term used to describe methods that change test conditions, test flow, test content and test limits (potentially at the die/unit or sub-die level) based on manufacturing test data and statistical data analysis. This includes feed-forward data from inline test and early test steps to later test steps and feed-back data from post-test statistical analysis that is used to optimize testing of future products.”

The key to adaptive test is to utilize data generated from testing the part or relevant data from previous processes or measurements in predicting the appropriate process for the future tests. Adaptive test needs to make decisions to reduce or increase testing as and when required. The ultimate goal is to apply only the minimum set of tests required to screen each part that will fail in the system either as shipped or over time [6].

## 1.2 Brief Description and Contribution of this Research

This thesis proposes a new Adaptive Test Scheme (ATS) to learn test fail rates *dynamically, per die* and *make real-time modifications to test order and test contents*. The thesis describes this adaptive test scheme that reorders and eliminates tests to reduce test times without compromising quality. Test elimination and reordering schemes need to estimate and model test fail rate of each test. Data already generated by the tester needs to be effectively utilized to provide statistical information for future test guidance. ATS utilizes Bayesian statistics to make dynamic updates to the test fail rates using per die pass/fail data for each test. ATS implements these updates using a Bayesian formulation that requires simple calculations which can be done in real-time on the tester.

Reordering tests according to test fail rates enables detection of failing die earlier in the test flow. This reduces the overall test time and test cost [7]. Adaptive test reordering has the following benefits:

- Reduces time to screen defects by yield learning and so reduces test cost
- Controls process and equipment variations at die, wafer, lot level.

- Provides opportunity to eliminate or sample test by sorting tests that are most likely to detect a fail and tests that are least likely to find a failing die.
- Reduces test time for a given product quality.
- Increases production capacity by testing more units for a given time.
- Accelerates production ramp-up by learning and adapting to product test data.

The Adaptive Test Scheme (ATS) described in this thesis demonstrates all the advantages listed above. In addition to all this ATS provides incorporates a simple natural monitor to control the impact of excursions, for example, unexpected incidence of a unique fail mechanism for an eliminated test.

Some part of this research has been presented at IEEE International Test Conference (ITC), in Sept. 2011, and published in the conference proceedings [8].

### **1.3 Organization of Thesis**

This thesis has been laid out in five chapters with different subsections. The next chapter describes previous work in the field of adaptive test. Further, it proposes a new classification scheme for the different adaptive test implementations and presents them in ITRS format. The third chapter of this thesis describes in detail the Adaptive Test Scheme (ATS), which enables real-time test reordering and elimination. The fourth chapter of this thesis a detailed evaluates ATS using synthesized data (Monte Carlo) and shows results of application of ATS to production data for two products from our industry partners. This thesis concludes by summarizing the benefits and requirements of the ATS developed in this work

and by making recommendations for researchers and product engineers working on adaptive test methods.

## Chapter 2

### Background: Previous work, New Classification and Motivation

Traditionally test flows are static, which means different tests and test patterns are applied in a predefined order. Most test application programs are written with the test application stopping at the first failing test referred to as stop-on-fail (SOF). A SOF test configuration gives lower test times for failing parts. Some parts are also subjected to continue-on-fail (COF) analysis where testing continues even after detecting a fail. COF testing is done by overriding SOF behavior and collecting pass/fail data for the full test set. The COF test data can be analyzed to make changes to the test program. Due to the large test data samples and long test times required by individual test effectiveness analysis, COF can be very costly [9]. So, COF based test effectiveness analysis is often not a practical way to adapt test programs as defect mechanisms do change through time.

Today, most manufacturers perform COF or SOF based volume test data (millions of units) fail analysis for a product and then provide recommendations to make changes to the test program. Examples of changes to the test program are: rearranging the test application order; modifications to the test content by eliminating or sampling test(selectively testing for few sample die); adjusting test limits, guardbands and setpoints. Test changes are then subjected to review by a change control board to check if there are any associated quality risks. Apart from being cumbersome and time-consuming, this process has the disadvantage of being unresponsive to short term temporal variation or spatial variation. Desire to control defects and variation has led to a lot of work in the field of test reordering,

test optimization, test pattern reduction, adaptive limit setting and adaptive test.

## 2.1 Previous work on adaptive test

The idea of adaptive test can be found as early as 1981 when a model for adaptive system diagnosis, but not test was described [10]. In 1991, test reordering using fail rates was first discussed by Huss and Gyurcik, where tests are reordered based on the pre-production estimates of faults and test fail rates, derived from a directed flow graph of all modeled faults in the circuit to be tested. This static reorder method works for small circuits with a few hundred transistors but becomes increasingly difficult to apply with increased circuit complexity [11]. Some papers on adaptive test present work on the problem of optimal test pattern ordering by using complete pass/fail information for all patterns based on a few sample die. An efficient heuristic to determine the pattern order to reduce the average test time for defective devices is proposed in [6]. The effect of the ordering of test types on test cost is analyzed in [7] using SEMATECH data. A method that models test escape rate as a function of multiple test coverages has been described in [12].

Adaptive test methods are primarily applied in the parametric test domain for recommending test elimination candidates and setting adaptive test limits [4–6]. The parametric tests like IDDQ, noise margin tests, specification tests, return a measurement and not just a pass or fail decision as in digital tests. The adaptive methods can perform correlation analysis based on parametric measurements from a sample units for each lot and adaptively select optimal parametric tests on a per lot basis. Part Average Testing (PAT) techniques involving calculation of process capability coefficients,  $C_p$  and  $C_{pk}$ , for the different parametric measurements to

decide test elimination candidates are now being adopted in the semiconductor industry [5].

The above methods require costly pass/fail information collection for the full test set and, most importantly, they are not adaptive in test order. Each method presumes that the initial test order remains effective throughout the wafer. Reordering tests in descending order of test fail rates detects the failing die earlier in the test flow and hence reduces the test time and test cost . Ideally, test order should adapt to variation in test fail rates caused by process variation and defects. A method using defect diagnosis to understand the changes in defect types throughout the production life-cycle, based on failure mechanisms is used to adaptively change test content in [13]. But a study involving defect diagnosis is time consuming and limits the frequency of adaptation to a few lots.

The 2009 International Technology Roadmap for Semiconductors (ITRS) predicted that by 2015 test pattern volumes will increase by  $25x$  causing  $11x$  more test times [3]. Ferhani in [14] suggests that 70% of the applied production patterns have very low unique fallout for digital circuits. Elimination of such patterns will have a little effect on DPPM but a relatively large impact on test time and test cost. Eliminating tests or test patterns with low fallout and testing few die with a reduced test content reduces the test time for good die. Any test elimination approach needs an accurate prediction of these low fallout tests and a careful study of DPPM impact.

The Adaptive Test Scheme (ATS) described in this thesis provides a method to adapt test order and test content to the test fail rate variation on a wafer and



builds on the concept of having an optimum test set (OTS) and order, for each unit tested as described in [6].

## 2.2 New classification of Adaptive Test techniques

Various adaptive test methods such as changing test limits to product data, performing outlier analysis to adapt test at the next test insertion etc. are being employed in the semiconductor industry. Some of these methods have been deployed on a large scale while others exist on an interim basis or smaller scale [5, 15, 16]. The key idea for all these adaptive methods is to manage process variation and achieve lower test cost at a better quality and higher yield. Broad definition and usage of the term adaptive test suggests a need for a general classification of these techniques to provide a clear roadmap for researchers and product/test engineers. Two classifications of adaptive test have been discussed in [3]. The first is based on when data analysis is performed.

- Real-time. In parallel with testing and on tester, per die.
- Near-time. At the end of sample testing, end of wafer or a lot or many lots.
- Off-line. Off-tester statistical calculations set test limits and control the test flow.

The second classification is based on whether adaptive test data is collected from a single or from multiple test insertions (wafer sort, burn-in, package test).

- Single Insertion. Data from single test insertion is used to adapt tests to local variation.

- Multi-insertion. Data from any test insertion is available to any other test insertion like wafer sort, package test, burn-in or final test as data feed-back and data feed-forward systems.

Combinations of when the data is collected and when it is analyzed are used at various insertions to set limits, change test flows and content, and modify setpoints.

This thesis presents an additional new third classifier for adaptive test based on different *Stages of Adaptive Test*. This classification is depicted in Figure 2.1. The figure is a modified version of table proposed in [17]. The color code used in this figure is the standard ITRS color code shown in Figure 2.2. Adaptive test classification is based on three stages of adaptation; simple, complex and continuous adaptation. A stage is defined by the attributes in each test method row. The remaining part of this section defines these terms and their attributes.

The adaptive test scheme described in this thesis is an example of a real-time, *Continuous Adaptation for Reordering and Eliminating tests* (marked in the Figure 2.2). The thesis demonstrates the use of this method at a single insertion using wafer sort test response data and takes a step towards continuous adaptation implementation.

Manufacturing test methods fall into three different categories and are defined as follows:

- Resetting Limits. Selecting the criteria for making pass/fail decision in case of parametric measurements.

Test Methods		Stages of Adaptive Test		
		Simple Adaptation	Complex Adaptation	Continuous Adaptation
<b>Single Insertion</b>	Setting Limits	S	◆ C	D
	Re-ordering Eliminating Tests	S	◆ C	○ D
	Changing Setpoints	S	C	C
<b>Multi-Insertion</b>	Setting Limits	S	◆ C	D
	Re-ordering Eliminating Tests	S	◆ C	D
	Changing Setpoints	-	◆ C	D
<i>S- Static sample independent (Decision taken: Offline)</i>		<i>C- Characterization For sample die (Near-Time)</i>		<i>D- Dynamic per die (Real Time)</i>

Figure 2.1: Classifying adaptive test in ITRS format. Adaptive test scheme described in this thesis is an example of a *Continuous Adaptation for Reordering and Eliminating Tests* encircled in the figure.

	Manufacturable solutions are known
◆	Interim solutions are known
	Manufacturable solutions are not known

Figure 2.2: ITRS standard color code.

- Re-ordering and Eliminating Tests. Process of changing the test flow and the test contents by reordering the tests to detect the failing die at the earliest test and eliminate certain tests or patterns that have lowest impact on DPPM.
- Changing Setpoints. The process of altering setpoints for test such as temperature, voltage and frequency based on data from previous die or other insertions.

The stages of adaptation fall in three different categories:

- Simple Adaptation. Test limits, test contents and test flow are statically set and periodically updated based on offline analysis, done by analyzing data acquired over long -intervals of time and volume such as after many wafer lots or mid production cycle.
- Complex Adaptation. Test limits, test contents and test flow are set by a characterization or sample set of die and are periodically modified based on near time analysis such as after looking at sample die on a wafer, a wafer or wafer-lot. The feed-back or feed-forward test data of these sample die from different test insertions may also be used to make changes to the test program. Most of the existing adaptive test techniques fall in this category. Examples are: Part Average Testing (PAT) uses data from previous sample die on a wafer to set the test limits for parametrics on-the-fly [5]; Reference Die Analysis scheme tests a sample set of die on the wafer with entire test suite and applies reduced testing to other die on the wafer by analyzing the test response data of reference die across multiple insertions [15]; Location Averaging technique screens statistical outliers using principal component

analysis to detect fails at downstream test insertions such as burn in and final test [16] . These techniques use few sample die to do statistical analyses near time or offline.

- Continuous Adaptation. Test limits, test contents, test flow and setpoints are adjusted continuously at single insertions as well as by using the data from different insertions, dynamically on a per die basis, real-time. Based on this classifier our adaptive test scheme supports continuous adaptation through reordering and elimination of tests. Our adaptive test scheme is demonstrated in a single-insertion mode but could take advantage of multi-insertions subject to availability of production feed-back and feed-forward data systems.

### 2.3 Motivation

Reordering by itself has no effect on quality as no test is removed from the test suite. The main objective of test reorder is to reduce the test time to fail. Wasted tester time and the tester depreciation cost associated with it is not the only economic impact of an inefficient tester program. ATE time that is needlessly spent testing a device could be better spent testing another device, which means that throughput is diminished when an inefficient tester program is used.

Test reordering using test fail rates needs a model to learn and update test fail rates and a rule which relates the test order update intervals and test fail rates. For example, if the fail rates are fixed and uniform then no test order updates are needed and tests can be in any order. If the observed fail rates are modeled to be slowly varying, then it suffices to use periodic quarterly or annual test order

updates based on fail rates obtained from large data samples. For rapidly varying test fail rates, frequent test order updates based on per-die, per-wafer or per-lot data may be needed.

ATS updates test fail rates with die pass/fail result employing Bayesian statistics to modify fail distributions and adopts a per-die test order update. Such a test reordering scheme leads to a significant reduction in variance of test time to fail for a DUT and hence the total test time. The reduction of test time and test time variance was a motivation for this research.

Figure 2.3 provides a preview of results to emphasize the variance reduction in test time to fail for a die on a wafer. The figure shows application of three test reorder and update models: Random Fixed Order, Characterization Test Order, and Adaptive Test Orders (from ATS). These results are for production test response data from 610 wafers and 16 tests of Product 1 described later in chapter 4. Figure 2.3 compares three stop-on-fail wafer test time distributions. The three distributions for test-time-to-fail are plotted for all failing die on 610 wafers of the product. The plot shows the potential for wafer total test time reduction using different test ordering schemes. The inner box-plot shows reduction in test time to fail variance for die for fails on a typical wafer from that data.

The adaptive test scheme (blue) is the die-level reordering scheme developed in this research. The adaptive test scheme has the lowest median wafer test time and the narrowest inter-quartile range of the three. The fixed test order (red) is any test order picked at random, called Random Test Order. As expected, the fixed order has a broadest inter-quartile range (IQR, used as a measure for variance

as it is immune to outliers) of the three and has the largest median wafer total test time. The Characterization Test Order (green) is the test order derived by sorting, in descending order, the test fail rates obtained from COF product data for a sample of 305 wafers. This order is then applied in a stop-on-fail (SOF) test flow to compute the green test time distribution shown in figure. This characterization order shows some improvement over the random fixed order. For example, the median and the IQR of the test time are smaller than the random test order. If test fail rates change rapidly with process variation and defect incidence, then only per-die test reordering using ATS can provide the benefits shown in this figure.

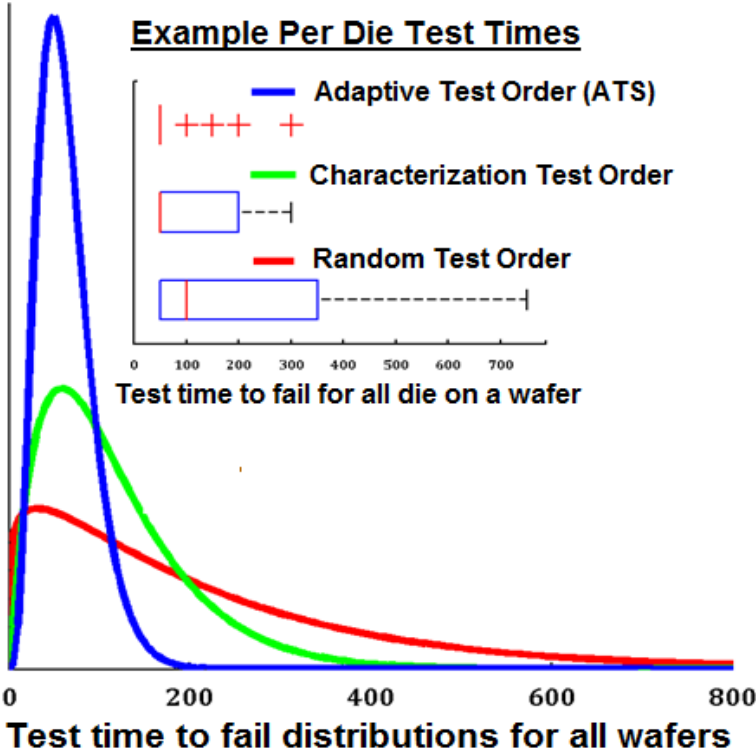


Figure 2.3: Test Time Variance Reduction for test-time-to-fail for three test orders. Each distribution represents test times for failing die on 610 wafers. The inner box-plot shows reduction in test time to fail IQR and outliers for failing die on a typical wafer from that data.

## Chapter 3

### New Method- Adaptive Test Scheme

#### 3.1 Using Bayesian Statistics

In Bayesian statistics, an educated guess about the probability distribution of the parameter(s) to be estimated is called a prior distribution. Next, the experiments designed to depend on the parameter(s) of interest are conducted and sample results or evidence are observed. Based on these observations, the initial guess about the distribution of the parameter(s) on interests is adjusted to reflect the current best estimate. This modified prior distribution called a posterior distribution.

The premise of Bayesian statistics is to incorporate prior knowledge, along with a given set of current observations, in order to make statistical inferences. The prior information could come from operational or observational data, from previous comparable experiments or from engineering knowledge. This type of analysis can be particularly useful when there is limited test data for a given design or failure mode but there is a strong prior understanding of the failure rate behavior for that design or mode. A posterior distribution summarizes the current state of knowledge about all the uncertain quantities (including unobservable parameters and also missing, latent, and unobserved potential data) in a Bayesian analysis. Posterior distribution(s) are obtained by incorporating prior information about the parameter(s) and inferences on the model parameters and their functions can be made [18]. Analytically, the posterior distribution is the product of the prior distribution and the likelihood.



The Bayesian update method used here is called the *Two-Stage Bayesian Method* described in [19]. A more specific discussion and application of this method where Bayesian statistics is used to predict equipment failure rate in reliability experiments can be found in [20]. To summarize the Bayesian approach consists of three main tasks:

1. Define a prior distribution for the variable to be estimated.
2. Collect evidence represented as a likelihood function.
3. Construct posterior distribution using Bayes' theorem.

Application of Bayes theorem for constructing a posterior distribution can be presented as a functional relation given in Equation (3.1)

$$f_{posterior}(\lambda) \propto P(X = x|\lambda) \cdot f_{prior}(\lambda) \quad (3.1)$$

where,  $\lambda$ = random variable to be estimated or modeled,  $f_{prior}(\lambda)$ = prior distribution of the random variable,  $f_{posterior}(\lambda)$ = posterior distribution of the random variable  $P(X = x|\lambda) =$  likelihood function as a function of  $\lambda$ .

## 3.2 Adaptive Test Scheme

### 3.2.1 Applying Bayesian Statistics for Learning Test Fail Rates

ATS uses Bayesian statistics to model and update test fail rate distributions for each test in the test flow. The Poisson probability model for test yield has been widely used in IC test for modeling random defects on a wafer and is an intuitive model for fail counts [21]. Assume a Poisson probability model for each test

in the test flow. The test fail rate  $\lambda$  is updated by observing fails on a wafer distributed with a Poisson likelihood. The choice of a Poisson distribution as likelihood function simplifies the calculations required to update distributions in Bayesian statistics and makes real-time computation possible. Assuming that each test has a separate fail rate  $\lambda$  the probability of a test having  $x$  failures is given by:

$$Pr(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \quad (3.2)$$

The test's yield would be given by  $Pr(X = 0|\lambda)$ . In presence of the variation and defects each test's fail rate  $\lambda$  is not a constant. The test fail rate  $\lambda$  varies across wafers and lots as discussed earlier in Section 2.2. In Bayesian statistics the conjugate prior of a Poisson likelihood function is a Gamma distribution  $\Gamma$ . ATS models the probability distribution of test fail rate  $\lambda$  of each test as a Gamma distribution [20]. These test fail rate distributions are the *prior distributions* and they represent best estimates of ATS about fail rates of those tests before testing a particular die on the wafer.

The Gamma distribution is described with two parameters; the first is a scale factor  $\alpha$  and the second a shape factor  $\beta$ . The Gamma prior distribution of a test is given by:

$$\Gamma_{prior}(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{(\alpha - 1)!} \lambda^{\alpha-1} e^{-\lambda\beta} \quad (3.3)$$

Application of Two-Stage Bayesian method to these prior distributions on occurrence of a pass or a fail gives the posterior test fail rate distribution for each test. The beauty of the (Poisson, Gamma) Bayesian formulation is that the *posterior distribution* for Gamma prior is a Gamma distribution  $\Gamma(\lambda|\alpha', \beta')$  where  $\alpha'$

and  $\beta'$  are updated based on the samples observed. In this model, updates to  $\alpha$  and  $\beta$  occur based on die test results given by,

$$\alpha' = \alpha + \text{number of fails} \tag{3.4}$$

$$\beta' = \beta + \text{number of die tested} \tag{3.5}$$

There is an  $(\alpha, \beta)$  pair for each test that gets updated as die are tested. The number of failing die screened by a test is represented by  $\alpha$ . The number of die for which that test was executed gives the sample size for that test represented by  $\beta$ . For each test executed on a die the fail rate distribution is updated by a unit increment to  $\beta$  and for each test failing a die fail rate distribution is updated by unit increment to  $\alpha$ .

This process of updating the test fail rates using Two-Stage Bayesian method can be depicted schematically as shown in Figure 3.1.

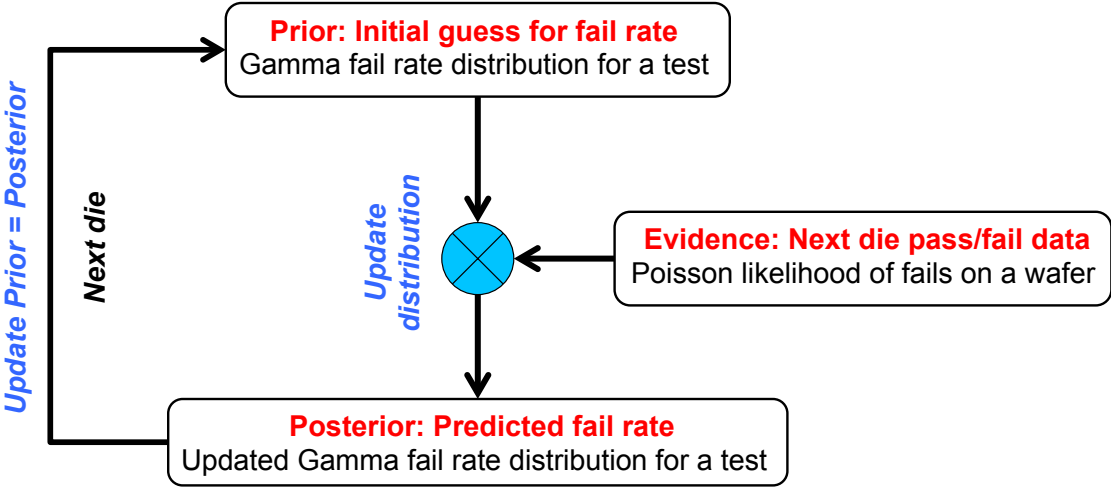


Figure 3.1: Flow diagram for Two-Stage Bayesian method in context of ATS.

Adaptive test flows and content are realized by comparing the statistics, such as means and variances of test fail rates, computed for each estimated posterior Gamma distributions to make per die decisions. The connection between Gamma distribution parameters  $\Gamma(\lambda|\alpha, \beta)$  and observed test fail rate statistics for each test can be established as follows

$$\text{Average, } E(\lambda) = \frac{\alpha}{\beta} \quad (3.6)$$

$$\text{Variance, } \sigma_{\lambda}^2 = \frac{\alpha}{\beta^2}. \quad (3.7)$$

Figure 3.2 shows a detailed flowchart describing the update strategy for ATS. The adaptive test scheme is composed of three loops denoted by loop 1, 2 and 3 in the figure. *Loop 3* is a loop over wafers and it defines the sample population or the number of wafers being tested. *Loop 2* is over individual die on the wafer, i.e. a step-by-step processing of the sample. *Loop 1* is over the individual tests giving the current best estimate of the test fail rate as described in Figure 3.1. When in the *Loop 1*, a test pass updates only the sample size  $\beta$ . If the die fails current test in the *Loop 1*, both  $\alpha$  and  $\beta$  are incremented and recorded.

On a failing die the test fail rate statistics are recomputed based on the new estimates of test fail rate given by updated posterior distributions. Based on the new statistics the test application order and content are modified. The new test order and content is then used for the next die to be tested in *Loop 2*. The posterior test fail rate distributions computed at the failing die are used as prior test fail rate distributions and are updated with the test results in *Loop 1*. For dies passing

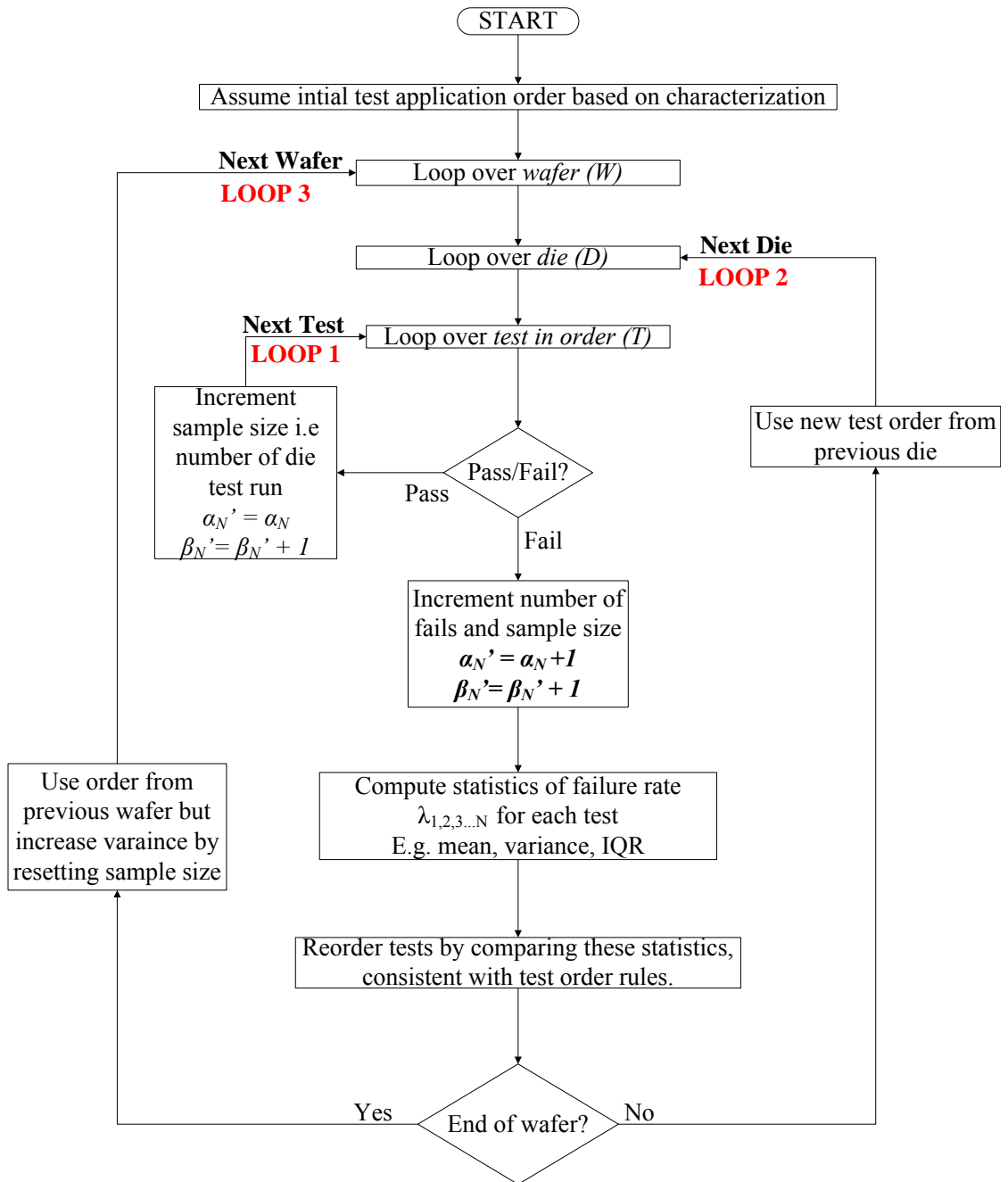


Figure 3.2: Flow diagram explaining Adaptive Test Scheme updates

all the tests,  $\alpha$  and  $\beta$  are not updated on any test. For dies failing a test,  $\alpha$  and  $\beta$  are not updated for tests following the failing test.

The initial test order can be any order, for example, an order obtained from fault simulation or characterization. For subsequent wafers the test order from the previous wafer is used but the sample size  $\beta$  is reset to diminish the bias of previous wafers test order. The previous wafers effect on the adaptation is limited by adjusting  $\alpha$  and  $\beta$  to preserve the average fail rate and increase the variance. This makes it easier for the scheme to respond to needed changes in test order on the new wafer [22].

Monte Carlo experiments described in Section 4.1 show that test order can adapt in response to local variation on wafer if the variances are increased at the start of each wafer. Hence per-wafer boundary for reset was chosen. Resetting sample size after a wafer provides a *natural monitor* in the adaptive test scheme. For an excursion forcing the scheme to choose a wrong test order the maximum latency for the scheme to react is a single wafer. After a wafer is tested completely, the sample sizes  $\beta$  are reset and the test orders are ready to adapt to local variations on next wafer. Monte Carlo experiments described in the next chapter will explain the effects of choosing different seed values, variances and the response of ATS on the impact of a shock. A shock is a sudden increase or decrease in fail rate of one or more tests.

Note that ATS is designed to update the test fail rates and test orders in a stop-on-fail (SOF) configuration, which inherently has censored data (no results for tests that were not applied). The ability of ATS to learn in the SOF configuration and

make correct decisions, shows it's importance to the success of adaptive test. If die are being tested in continue-on-fail (COF) then ATS flow can be easily modified to keep testing and updating test fail rates even after a die fails. The additional data available will make ATS estimate test fail rate distributions even more accurately. The evaluation of ATS provided in this thesis is for testing die in SOF configuration for two main reasons: 1) Test ordering with censored data is the harder problem to solve; 2) Most digital products like microprocessors, DSPs and microcontroller are tested in a SOF configuration [3].

Reordering has to be consistent with *test order rules* related to stress conditions, voltage and temperature set points, continuity tests etc.. For example while re-ordering pre- and post- stress tests cannot be mixed. While reordering tests, ATS has to treat different setpoints such as different voltages ( $V_{max}$  and  $V_{min}$ ) or temperatures as a group of tests and reorder tests for the same setpoint. This fail rate update strategy can be applied to any tests generating a pass/fail result regardless of the type of measurement performed. The method can also be applied to other test insertions for example final test and burn in.

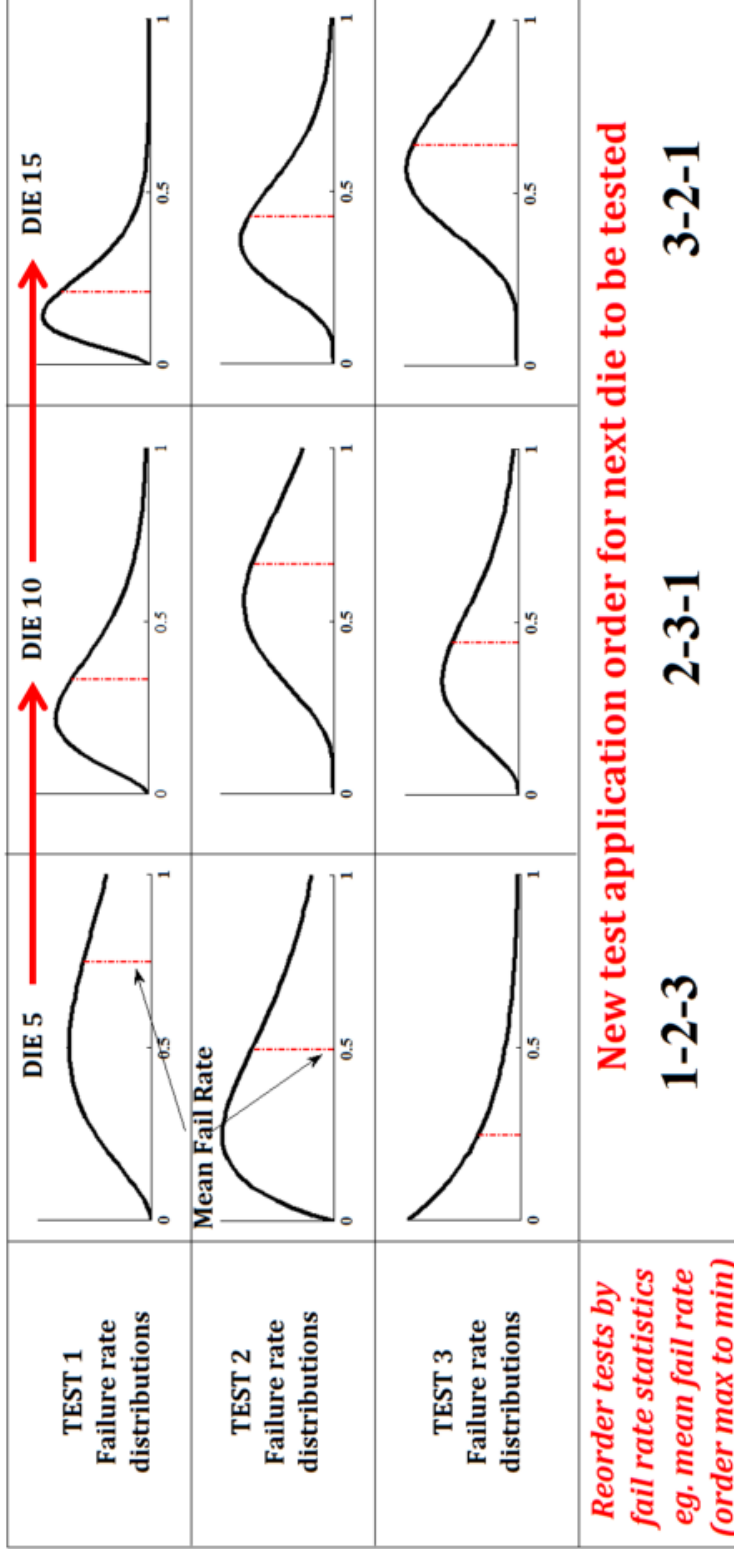


Figure 3.3: Adaptive test scheme application to die on a wafer showing updates to fail rates of each test and test reordering.



Figure 3.3 shows a simple example of the changes in the test fail rate distributions while testing die on a wafer. ATS estimates of the test fail rate distributions are updated at each failing die. For the first die on the wafer ATS assumes a prior distribution and updates the prior on each failing die. The  $x$ -axis plots the fail rate  $\lambda$  and the  $y$ -axis plots the probability of having that fail rate, given by a Gamma distribution. Each column gives the snapshot of estimated posterior Gamma distributions of the fail rates for tests 1, 2 and 3 at 5th, 10th and 15th die on the wafer. The statistics of these posterior fail rate distributions are used to reorder the tests. The Gamma distributions of failure rate of each test are updated along a row in the figure, depending on die failing that test. The five die interval is selected to display visible difference in the distributions, but parameter updates take place on each failing die. Test reordering is done by computing and sorting the statistics of the fail rate of each test, for example, the mean fail rate. After each die is tested, tests are reordered in descending mean-fail-rate order.

### 3.2.2 Test and Test Pattern Elimination

ATS reorders the tests in descending order of their fail rates by estimating test fail rate distribution for each die. Tests or patterns appearing later in the test application order are less likely to fail a die on the wafer than tests or patterns which are ordered to be earlier in the test application order. Test reordering by test fail rates reduces test time to fail for failing die. In chapter 4 it is shown that only way to reduce test time for good (passing) die is to test them with reduced number of tests or patterns i.e. eliminate tests or patterns from the test flow. ATS reorders the test in a way to reduce the test-time-to-fail variance and so the majority of fails appear at the start of test flow. The tests at the end of the test application

order are the most likely candidates for elimination after assessing the associated risk. A Monte Carlo method for understanding the risk of test elimination has been developed and is described in Chapter 4.

Any scheme of test elimination needs to have a monitor that limits the impact of an excursion in the process. An example is an unexpected incidence of a unique fail mechanism for an eliminated test. Examples of monitors include sensors on the silicon (e.g. ring oscillator structures, E-test structures) or sample testing some die in a continue-on-fail (COF) fashion. These monitors are expensive and they affect the test time reduction benefit achieved through such elimination schemes. ATS provides a *natural monitor* for reducing the impact of excursions without any additional costs. At each new wafer, ATS starts all over again with all the tests in the test program regardless of tests eliminated on previous wafer. While doing so, ATS preserves and starts testing the next wafer with the test order obtained from the previous wafer and resets sample sizes of test fail rate distributions. Thus, ATS monitors excursions by detecting unexpected changes in test order as shown in next section. The worst case latency for recovering from an excursion in this case is the test time to test one wafer.

### **3.3 Adaptive Test Scheme - Monte Carlo Experiments**

Monte Carlo is a popular computational technique used for evaluating risk in quantitative analysis and decision making under uncertainty. Most Monte Carlo experiments are set up in the following four steps:

1. Define a domain of possible inputs.
2. Generate inputs randomly from a probability distribution over the domain .

3. Perform a deterministic computation on the inputs.
4. Observe and aggregate the results by repeating the process multiple times.

Monte Carlo experiments, when set up as described above, provide useful information about uncertainty of the inputs, sensitivity of a computation technique to different seed values and risk associated with decisions made. Monte Carlo experiments simulate a variety of outcomes for different inputs and help to assess the impact of choices made as well as allow us to understand the technique better. Hence, these experiments are of great value while designing adaptive test methods.

ATS makes decisions to dynamically reorder tests on a per die basis. Monte Carlo experiments can be used to check if these reorder decisions are correct by generating synthetic test data with known test fail rates and feeding the data to ATS. Bayesian inferences depend on the choice of a prior distribution. The seed parameters of these distributions influence Bayesian learning schemes and ATS is no exception. So, Monte Carlo analysis is useful for studying the effect of these seed parameters and making good seed choices. Chapter 4 shows an analysis of the effect of choice of seed parameters on ATS predictions for test flow changes and underlying test fail rates.

To reduce test time for good dies, ATS identifies tests which can be eliminated from the test flow . Test elimination involves many decisions such as choosing appropriate candidates for elimination, triggering elimination at some point on the wafer and number of tests to be truncated. Monte Carlo simulations provide a tool to analyze these different scenarios and assess the impact of choices made. Chapter 4 shows a Monte Carlo simulation example to determine the effect of

choice of a truncation trigger.

The different steps describing the flow for the Monte Carlo simulator are listed as follows:

1. Input the number of die on a wafer and the number of tests for which pass/fail data is to be generated.
2. Input test fail rate Gamma distributions for each test to generate test pass/fail data for entire wafer by selecting  $\beta$  and average fail rate  $\lambda$  for each test.
3. Pick random samples of the test fail rate  $\lambda$  from test fail rate distributions and derive die pass/fail results by substituting  $\lambda$  in equation 3.2 to calculate Poisson probability for  $x = 0$ .
4. Apply ATS to this synthetic wafer and record the different ATS test orders.
5. Repeat generation of synthetic wafer test data and application of ATS to this wafer multiple times.

Results of the Monte Carlo experiments are described in Chapter 4. These results show that ATS learns the average order of test fail rates of different tests on a wafer.

## Chapter 4

### ATS Performance Evaluation: Monte Carlo and Production data

#### 4.1 Monte Carlo Evaluation of ATS

##### 4.1.1 ATS Test reordering for synthetic test data

The Monte Carlo simulator can be used to study the response of ATS to a *shock* increase in fail rate of one or more tests. Synthetic data generated using known underlying test fail rate distributions, provide the ability to verify the working of ATS to order the tests correctly by learning these fail rate distributions on the wafer and adapting to shocks. One example application of ATS to die on a synthetic wafer is shown in Figure 4.2. Test fail rate distributions used for generating data for the synthetic wafer are shown in Figure 4.1. The data for die on the wafer before die 270 is generated using test fail rate distributions T1, T2, T3 (before shock), T4 and T5 and data for die on wafer after that is generated using test fail rate distribution from T1, T2, T3 (after shock), T4 and T5.

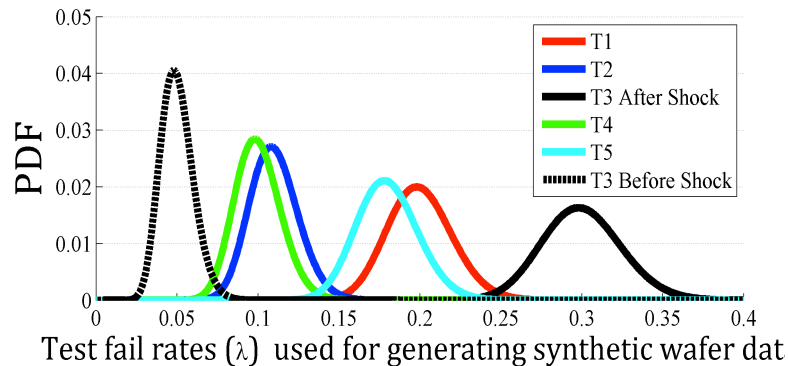


Figure 4.1: Test fail rate distributions ( $\lambda$ ) for five tests used to generate synthetic wafer test data after shock for test fail rate of Test T3.

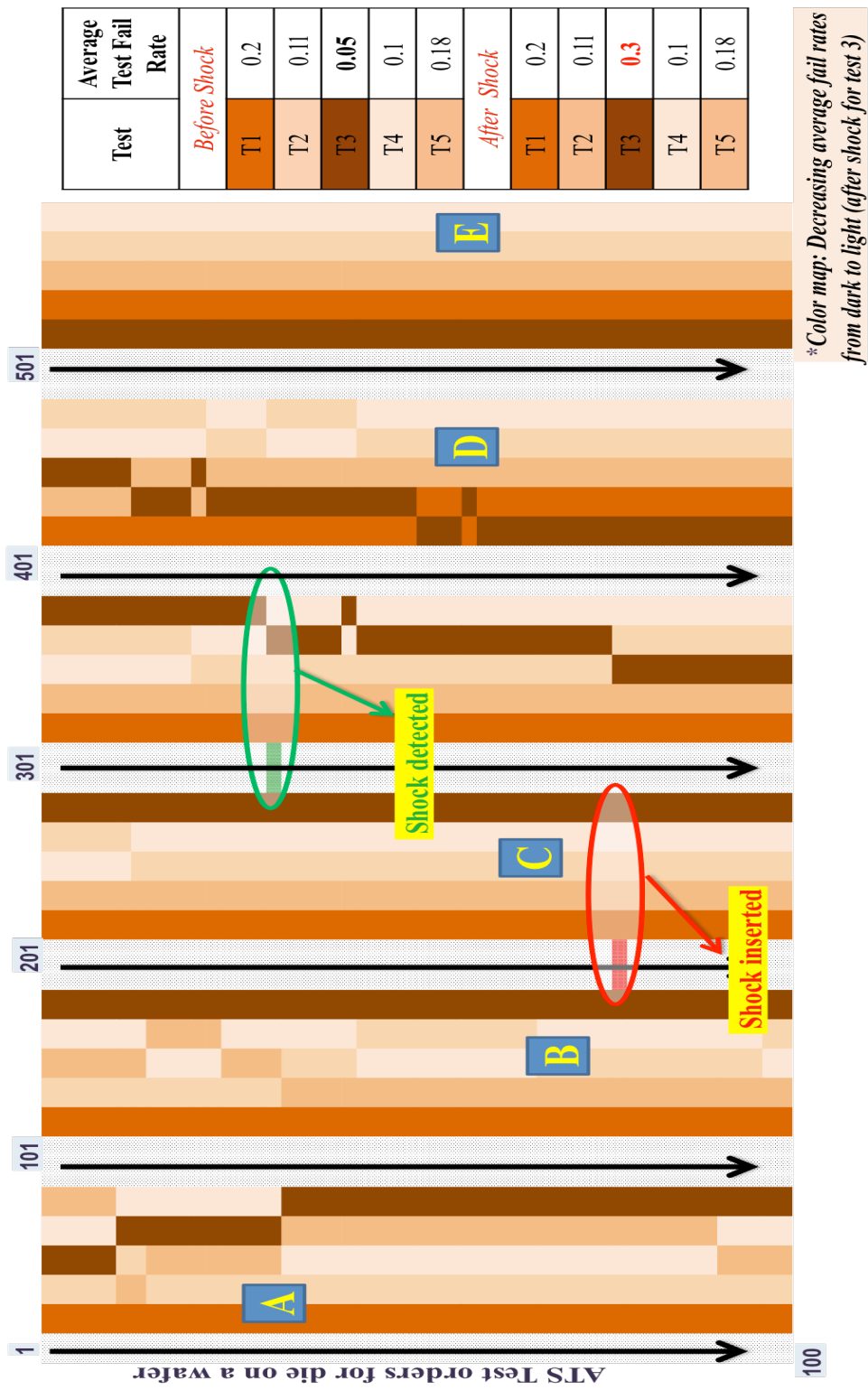


Figure 4.2: ATS test order for a wafer of 600 synthetic dies tested with 5 tests, divided in groups of 100 for convenience. Initial test order settles into an order driven by *before shock* test fail rates. A shock occurs at die 270 causing T3 to have the highest fail rate.

Figure 4.2 shows the adaptive test orders at different die locations on the synthetic wafer with a 50% yield in form of a color map. The darkest color is assigned to the test with highest fail rate on the wafer and lightest color is assigned to the test with lowest fail rate, after the shock. Note that the ATS is applied in a SOF configuration as described earlier. This figure is divided into 5 regions *A*, *B*, *C*, *D* and *E* and the test order updates in each region can be explained as follows:

- Region A. ATS starts testing the wafer with the test order T1-T2-T3-T4-T5 on the first die. Test order updates in this region show ATS working to learn test fail rates of the test and updating the estimates of test fail rate at the failing die. Observe how quickly the lowest average fail rate test (before shock), T3 is pushed to the end of the test flow.
- Region B and C. Region B and C, show ATS settling down on the correct average test order for the tests T1-T5-T2-T4-T3 before shock.
- Shock insertion and Shock detection. The test fail rate for test T3 is subjected to a shock increase, bumping it up from 0.05 to 0.3 at die 270. ATS detects this shock after 63 die at die number 333 by changing the test order to move test T3 up in the test flow. Note the busy portion of the wafer after shock detection adapting the test orders to the shock increase in fail rate.
- Region D and E. In region D on the wafer, after a series of updates to the test fail rates and the test orders ATS moves test T3 to the top of the test flow and settles on a new order. In region E, observe that ATS settles on a test order which us the average test order of the test fail rates used for generating synthetic test data.

### 4.1.2 Effect of choice of seed variance on ATS

Bayesian modeling requires a few seed parameters. The initial seed parameters are initial guesses to start the algorithm at the first die of the wafer to be tested. Different seed values have a considerable influence on the Bayesian learning. The adaptive test scheme (ATS) is no exception. Monte Carlo simulations are used to characterize seed value performance and thereby select good seed values. For example, a key ATS parameter (seed) is the initial guess for the variance of each test fail rate.

This section gives results of the application of ATS with narrow and wide variance seeds to synthetic test data generated using a Monte Carlo simulator. The variance of a Gamma distribution  $\Gamma(\alpha, \beta)$  given in Equation 3.7, varies inversely with the square of shape parameter  $\beta$ . Note that this seed variance value is used only for the first die of the wafer to be tested. Every new wafer is seeded by test order and average fail rates learned from the previous wafer, but variances are reset at the new wafer to the seed value. Resetting variance allows ATS to adapt to local variation and defects on the new wafer.

Each Monte Carlo experiment starts with generation of synthetic wafer test data from known test fail rate distributions. The Monte Carlo simulator can be configured to generate test response data for a any number of tests and die on a wafer. All the Monte Carlo experiments described in this thesis generate a synthetic wafer with 5 tests and test response data for 700 die. Keeping the number of tests low makes it easier to follow the interaction of test fail rates and ATS reordering, while presenting the results effectively.



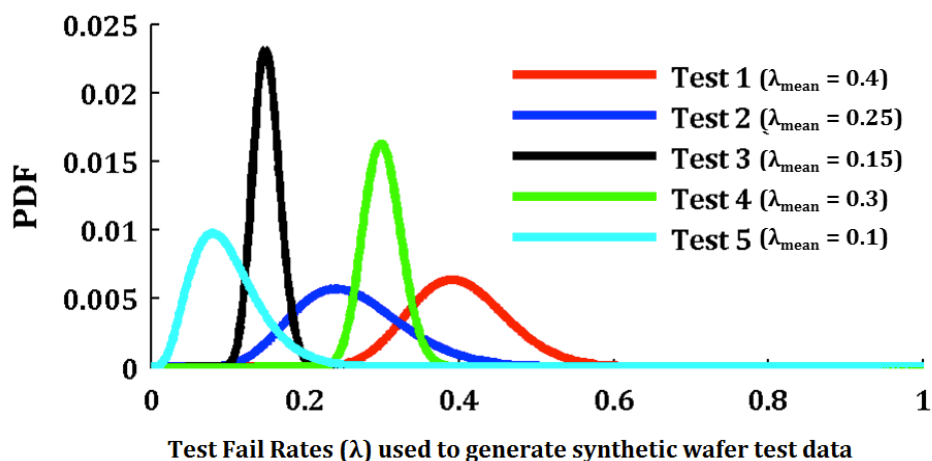


Figure 4.3: Five synthetic test fail rate distributions used in Monte Carlo testing of the adaptive test scheme. Each fail rate is modeled as a Gamma distribution with its own mean and variance. Test fail rate means are shown in the legend.

The effect of choice of seed variance on ATS is explained by using an example Monte Carlo run. Note that apart from this example, different test fail rate distributions using different average fail rates and shapes like bell, right tailed, left tailed etc. were considered and the performance of ATS was assessed. The following example conveys results and knowledge gained from all these experiments. For this example, the five synthetic test fail rate distributions used to generate test data using Monte Carlo simulator are plotted in Figure 4.3. These five distributions have different means and variances.

ATS is applied to synthetic Monte Carlo wafer test response data and estimates for the test fail rates and the distributions are obtained. This synthetic generation of test data and application of ATS was repeated twenty times. Figures 4.4 and Figure 4.5 summarize the results of these experiments by presenting a typical result for use of a narrow variance seed and a wide variance seed. Seed distributions used

for this analysis had equal mean fail rates.

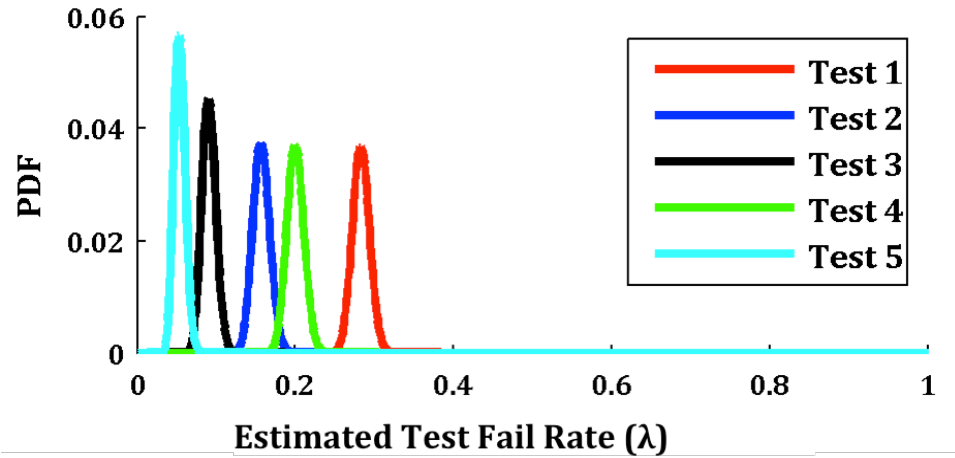


Figure 4.4: Computed fail rate distributions after application of adaptive test scheme (ATS) for a narrow variance seed. ATS finds the original order of average test fail rates on the wafer

Figure 4.4 shows the distribution of test fail rates estimated by ATS at the last failing die on the wafer using a narrow variance seed for the initial guess at the first die. Starting with the narrow variance seed, the adaptive test scheme underestimated the means of every test. However, ATS ordered the tests correctly. Figure 4.5 shows the results starting with a wide variance seed. With the wide variance ATS estimates the five test fail rate means to be nearly equal. As a result, ATS test ordering did not reflect the underlying structure of the fail rates in Figure 4.3 and wafer test times were much larger and similar to the random test order. The difference in ATS test order and total wafer test time as a function of narrow and wide variance was confirmed with 200 synthetic wafers, with 10 different test fail rate configurations (repeated 20 times each). This analysis shows that ATS guesses the correct order of average fail rates of the underlying distributions when

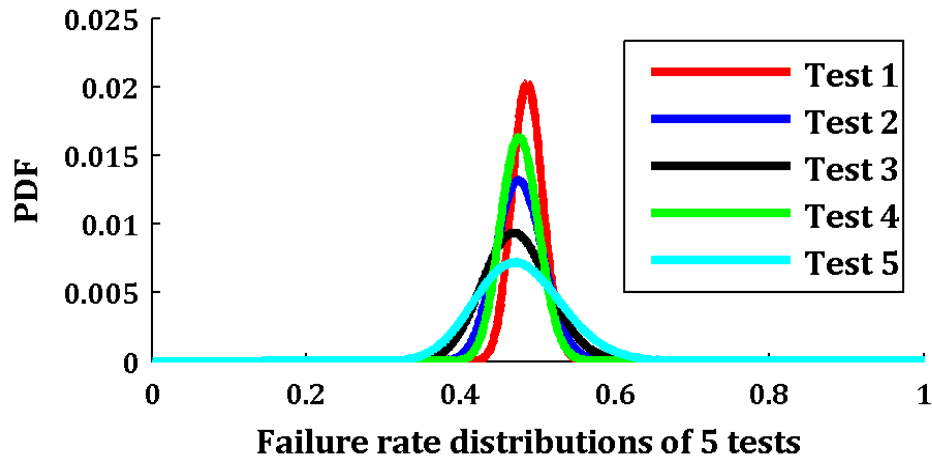


Figure 4.5: Computed fail rate distributions after application of ATS for a wide variance seed. ATS ordered the five tests at random with test times similar to the random test order.

initialized with a narrow variance distribution. The narrow variance seeds were used in all further applications of ATS.

#### 4.1.3 Effect of choice of test elimination trigger on ATS

Two choices need to be made while eliminating tests: 1) when to start or trigger test truncation and 2) how much testing to truncate. Both these choices are directly related to the quality level of the product and carry a risk of an excursion such as the sudden shock increase in fail rate of a test that is selected to be eliminated. The second choice of how much to truncate or how many tests to eliminate from the test content depends on the test time reduction (TTR) and quality (in terms of DPPM) targets for the respective products. Analysis of this tradeoff is shown in Section 4.2.3. The first choice of when to trigger truncation on wafer needs some deliberation and is discussed in this section.

Test elimination using ATS makes use of ATS's ability to estimate the average test fail rate for a wafer and settle down on a test order derived from that test fail rate. Furthermore, a high correlation can be observed between consecutive test orders for the tests ordered to be at the end of the test flow . This high correlation event will be shown in (see Figure 4.11) using real production data and will be explained in Section 4.2. ATS makes the choice of selecting the tests at the end of the test order with lowest estimated test fail rates as likely candidates for elimination. ATS updates the test fail rate distributions on failing die and makes a reordering decision based on these fail rate distributions. The choice of when to truncate depends on the number of fails required by ATS to learn the average test fail rate on a wafer, and settle down to a test order which has high correlation between consecutive test orders for tests at the end of the test order.

This settling down of ATS test order can happen after a few fails at the start of the wafer or after testing quarter or half of the wafer depending on the yield of the product, the number of die on a wafer and the number of tests or test patterns under consideration. There is no one magic number of fails after which truncation can be triggered on a wafer for all the products. However, for individual products information about the yield, number of die on a wafer, number of tests can be used to assess the impact of shock increase in fail rate of eliminated test. Specific information can be used by the Monte Carlo simulator and impact of events such as shock before or after truncation trigger can be analyzed. The objective of this section is to give an example to show how Monte Carlo simulator can be used for such an analysis.

For this example, five tests with different fail rate distributions were used. The average fail rates of the five tests are given by Table 4.1. These test fail rates generate a typical wafer with 700 die and 350-400 fails and about 50% yield. As an example the test truncation trigger is selected at 160th fail on the wafer and only one test is to be truncated or eliminated. A test fail rate shock is modeled by generating test data from a distribution with higher test fail rate for all die on the wafer after the insertion of the shock. In this example shocks are inserted at fails ranging from 50, 80, 100, 120, 140, 160, 180 and 200. The shock is inserted for test T3 whose average fail rate is increased from 0.01 to 0.3 as shown in the Table 4.1. In normal operation of ATS, with no shock in test data, ATS will eliminate test T3 because T3 has the lowest fail rate.

Table 4.1: Example test fail rates for demonstrating use of Monte Carlo simulation tool for studying effect of shock on choice for truncation trigger.

<b>Test</b>	<b>T1</b>	<b>T2</b>	<b>T3</b>	<b>T4</b>	<b>T5</b>
Average test fail rate <i>Before Shock</i>	0.2	0.1	0.01	0.15	0.18
Average test fail rate <i>After Shock</i>	0.2	0.1	<b>0.3</b>	0.15	0.18

This process of generating synthetic wafers with shock inserted at different fails on the wafer described above and applying ATS to these wafers was repeated thirty times. Figure 4.6 shows a box plot for the test escapes per wafer after application of ATS to these wafers. The x-axis represents shock insertion at different failing die. The truncation trigger is fixed at 160th fail on the wafer. Observe that the test escapes reach a peak when shock is inserted at 160th fail which is the same point when the truncation is triggered, giving ATS no time to learn and react to the shock. Similarly, if the shock is inserted at fail number 180 or 200, ATS would

have already truncated test T3 at the 160th fail and hence would not detect the shock. Even if the shock is inserted at 140th fail ATS is not able to increase fail rate estimate for test T3 to be at the start of the test order and ends up eliminating it.

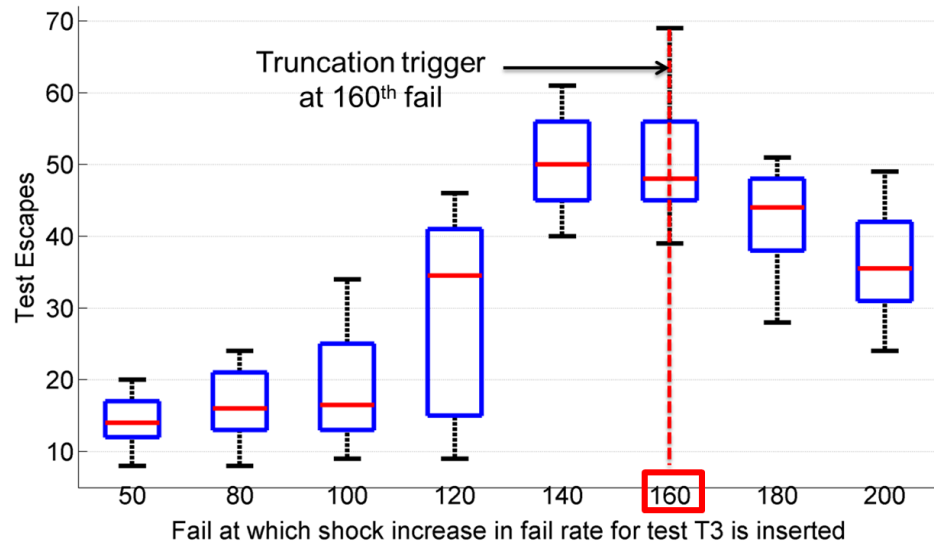


Figure 4.6: Effect of shock inserted at different failing die number on test escapes when truncation trigger is set at the 160th failing die.

Table 4.2 shows number of wafers for which ATS test eliminated tests T2, T4 or T3 for the different shock insertions. Note that in this experiment one test has to be eliminated for every wafer and this table captures which one got eliminated. Observe that for shock inserted at fail number 50, ATS detected the shock increase in fail rate before it tested the truncation trigger die. The test, the lowest average fail rate test after the shock, was eliminated on all 30 wafers. When shock was inserted at fail 100 and 120, ATS detected the test T3 shock insertion and did not eliminate it. However, for shock insertion at fail 140 and 160 ATS could not detect

the shock and ended up eliminating test T3 on all the 30 wafers.

This simplified example is intended to show how ATS works, and should not be considered as a specific recommendation for the number of fails or as a measure of the sensitivity of ATS to test shocks. This is just an example which shows truncation in a scenario of only 5 tests, all having high fail rates and a low yielding wafer to make it easier to comprehend. The example demonstrates how Monte Carlo can be used to analyze the characteristics of ATS.

Table 4.2: Number of trials having different tests eliminated for shocks inserted at different points on synthetic wafer

<b>Shock inserted at fail number</b>	<b>T2</b>	<b>T4</b>	<b>T3</b>
50	30	0	0
80	30	0	0
100	28	2	0
120	20	3	7
140	0	0	30
160	0	0	30
180	0	0	30

## 4.2 Application to Production data

### 4.2.1 Description of Production Data

Wafer sort test response data was obtained from a leading semiconductor chip manufacturer for two integrated circuit products. The production data was used to demonstrate the test time benefits of ATS test reorder and elimination. The test data for both the devices was collected in a continue-on-fail (COF) configuration (see Chapter 2) with test times for each test. Since test results for all the tests and all die were available, the test flow could be emulated in stop-on-fail (SOF)

configuration using different test orders. SOF emulation of a COF dataset also allows computation of quality loss in terms of defect level (DPPM) after eliminating tests and patterns using ATS.

Product 1 is a 90nm integrated circuit chip comprised of a DSP core, memory and other logic. COF test data for 419k units of this product was analyzed. This dataset contains 610 wafers divided in 26 lots with 688 die on each wafer. The test data used contains test results for 16 digital tests like transition fault tests and patterns, stuck-at tests and patterns, Built-in-Self-Test (BIST) results for different parts of the chip. All these tests were assumed to have same unit test time. Test results for executing the same tests before and after stress (pre- and post- stress), at different voltages ( $V_{max}/V_{min}$ ) and using different capture mechanisms (launch on shift/launch on capture) for this device were processed through the ATS flow. Test reordering was done consistent to test order rules as described in Chapter 3.

Product 2 is a 65nm mixed-signal wireless integrated circuit chip. COF test response data for 1.2 million units of this product was analyzed. ATS application to this product was limited to test pattern reordering as test times for these test patterns were available. This dataset contains 176 wafers divided in 13 lots with approximately 6600 die per wafer. Test responses and test times for different transition fault patterns of this product were considered for the experiment. These transition fault patterns are graded according to fault coverage and divided into twenty almost equal parts referred to as test data loads (TDLs). ATS was applied to reorder and eliminate some of the 20 patterns. The initial test application order for these TDLs was obtained by ordering fault coverages and test times for individual patterns.



### 4.2.2 Reordering Tests and Patterns

ATS was applied to data from Product 1 and Product 2 in a SOF configuration. Test reorder does not affect test time to pass for a good die as a passing die has to go through the entire test suite. But, re-ordering reduces test-time-to-fail for a failing die and hence reduction in test-time-to-fail is used as a metric to measure success. Figures 4.7 and 4.8 show the histogram of percent test-time-to-fail reduction for all wafers in the both the products. In each figure the  $x$ -axis represents the percentage test-time-to-fail reduction per wafer. The  $y$ -axis represents the number of wafers which had  $x\%$  test time to fail reduction. The percentage test time to fail reduction per wafer is calculated with respect to the original default test order for these units.

Figure 4.7 shows that for Product 1 average ATS test time to fail reduction per wafer was 56%. For any given wafer in the dataset the test time of failing die was reduced by half. The test-time-to-fail reduction for some wafers is above 90% while for some it is below 10%. This phenomenon can be observed because test-time-to-fail reduction per wafer is dependent on number of fails on that wafer because the ATS uses the many fails to learn and update test fail rates. For high yielding wafers, the benefit of test-time-to-fail reduction is limited because failing units for which test time reduction (TTR) can take place i.e. fails are fewer in number.

Figure 4.8 shows ATS test-time-to-fail reduction for transition-fault-patterns in Product 2. Observe that average test time to fail reduction per wafer is an impressive 33%. This reduction in test time to fail is again dependent on yield of wafer for this product. Since only transition-fault-patterns were considered, ATS has fewer opportunities to update the estimates for test fail rates compared to Product 1. For both examples in this section, ATS is achieving TTR by reordering only.

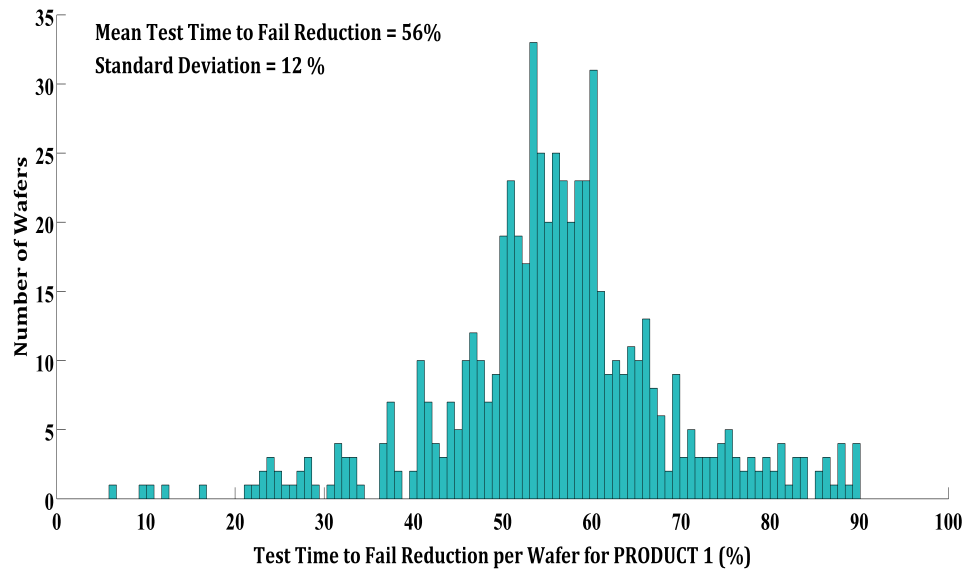


Figure 4.7: Test-time-to-fail reduction (failing die only), per wafer for Product 1 after application of ATS with reorder only.

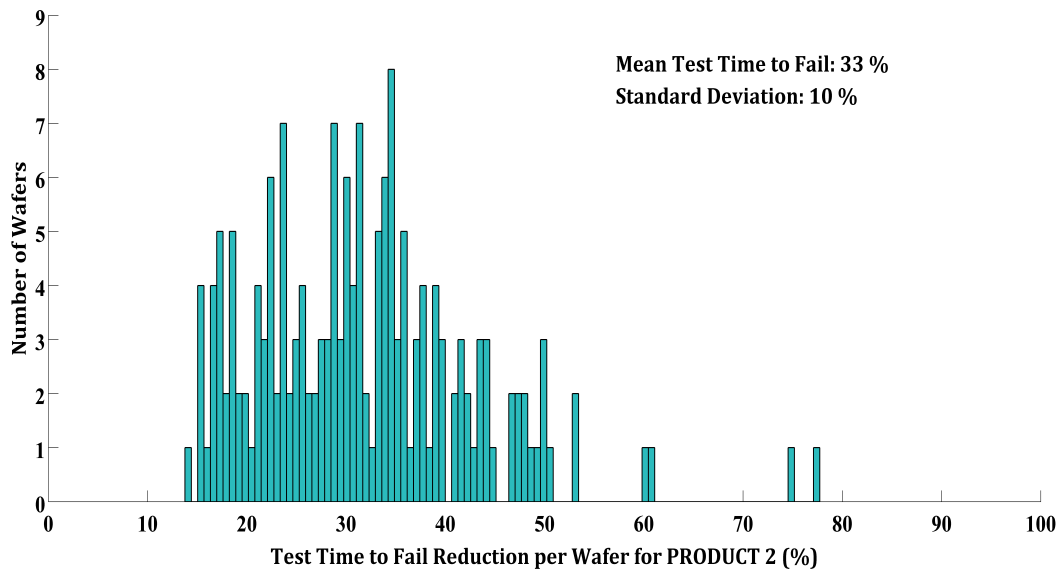


Figure 4.8: Test-time-to-fail reduction (failing die only), per wafer for Product 2 after application of ATS with reorder only.

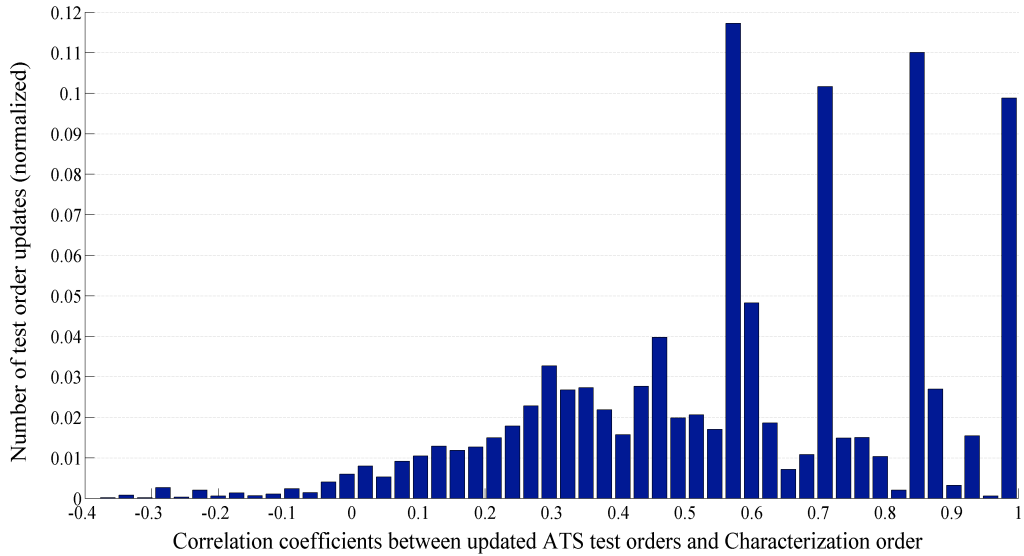


Figure 4.9: Correlation between ATS test orders and characterization order for test data from Product 1.

Figure 4.9 shows the histogram of correlation coefficient between different updated ATS test orders and the fixed characterization test order for all wafers of Product 1. The characterization order is the test order derived from analysis of COF test data for 305 wafers using obtained test fail rates. The characterization test order is used as default test order in test program. Note that only 10% of ATS orders correlate perfectly (i.e. = 1) with characterization order. The remaining 90% orders enable ATS to achieve reduced test times and adapt test orders to the test fail rates on a per die basis. In Figure 4.7 the three peaks apart from the peak for perfect correlation with characterization order, represent six different test orders. These six orders are the six orders for which ATS settled the most for all the wafers of Product 1. This shows that the characterization order i.e. test order used by the test engineers in the test program was *not* the optimum order.

ATS without test elimination, provides significant amount of reduction in test-time-to-fail per wafer, but this overall test time reduction (TTR) is diluted by the yield of the wafer. Reordering can achieve more TTR benefits for lower yielding products than products having very high yields. In general, products with smaller die size and very high yields will have lower TTR benefits by ATS reorder only, than products with larger die size and lower yields. Passing die (or good die), continue having long test times as they are tested by all tests in the test suite. The only way to reduce test time for these good die is by reducing the test content. ATS reduces test time of these passing die by eliminating tests ordered to be at the end of test flow by reordering, as described in the next section.

Effect of yield on TTR obtained by reordering test using ATS can be studied by simulating fails in production test data of Product 1 for different hypothetical yields shown in Figure 4.10. Observe the dilution of total TTR per wafer by yield. Product with higher yield (assumed 95%) has lowest average TTR of 6% , while the one with lowest yield (assumed 20%) is able to achieve highest average TTR of about 47% on application of ATS. Note that these TTR benefits are obtained by only reordering the tests.

The effect of yield on test time reduction can be explained with the help of following equation:

$$TT_{total} = Y \cdot TTG + (1 - Y) \cdot TTF \quad (4.1)$$

where  $TT_{total}$  = Total test time for a wafer,  $TTG$  = Total test time for good (passing) die,  $TTF$  = Total test time for failing die,  $Y$  = Yield of the wafer.

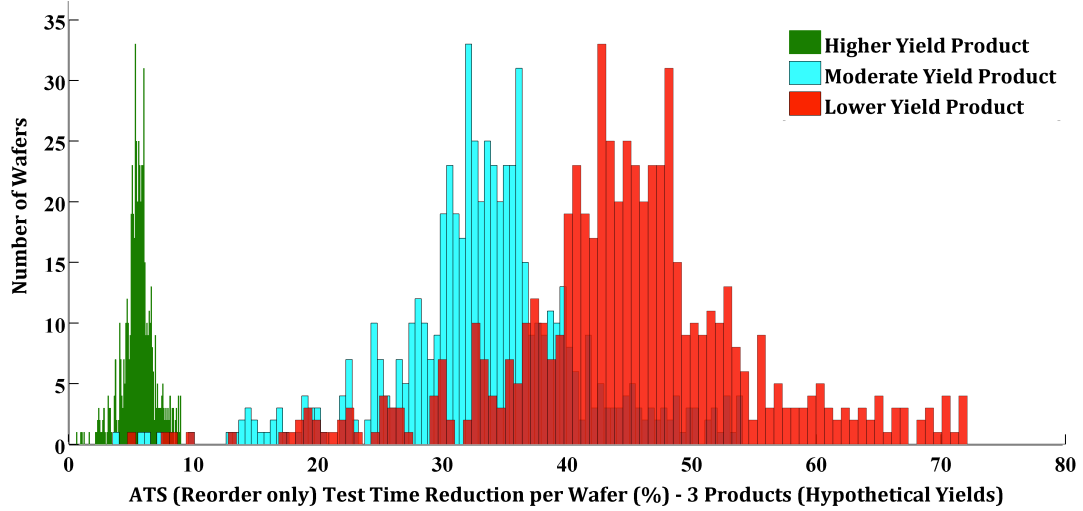


Figure 4.10: Results of application of ATS (reorder only) to three products with hypothetical yields, for the fails in Product 1.

Substituting values in the equation for a product with wafer yield 90%, reduction in TTF by 50%, reduces total test time by 5%. These estimates suggest that significant reduction in total test time can be achieved only when test time to pass is reduced. Eliminating patterns and tests from the tests suite is the only way to reduce test time for passing die. The results for eliminating tests with ATS are discussed in the next section.

#### 4.2.3 Eliminating Tests and Patterns

The ATS reorders the tests in descending order of their estimated average fail rates across the wafer. Tests reordered to be at the end are most likely candidates for elimination. After a certain number of fails, the tests which ATS ordered to be at the end of the test order did not change their position much or remained stable. Test reordering mainly took place only between tests at the start of the test application order and the rest of the order remained stable suggesting that

ATS settled down on a test order for the tail of the test flow.

This can be shown by plotting a histogram of correlation coefficient between consecutive test orders for die on the wafer shown in Figure 4.11. A high correlation can be observed between the two orders as most samples tend to populate the tail end of this chart. This arises from the fact that even if changes take place in the test order after a short settling time, these changes are only characterized by shuffling of the high fail rate tests at the start of the test order. The tests late in the test order can be removed without having much impact on the DPPM. Note that there exist orders which have negative correlation coefficient which suggests that some tests had a significant change in their position in the test flow. This phenomenon is observed at the start of the wafer where the order changes and adapts very quickly to learn the fail rates. As the adaptive scheme learns the fail rate distributions these order reversals do not take place and a more stable order is established.

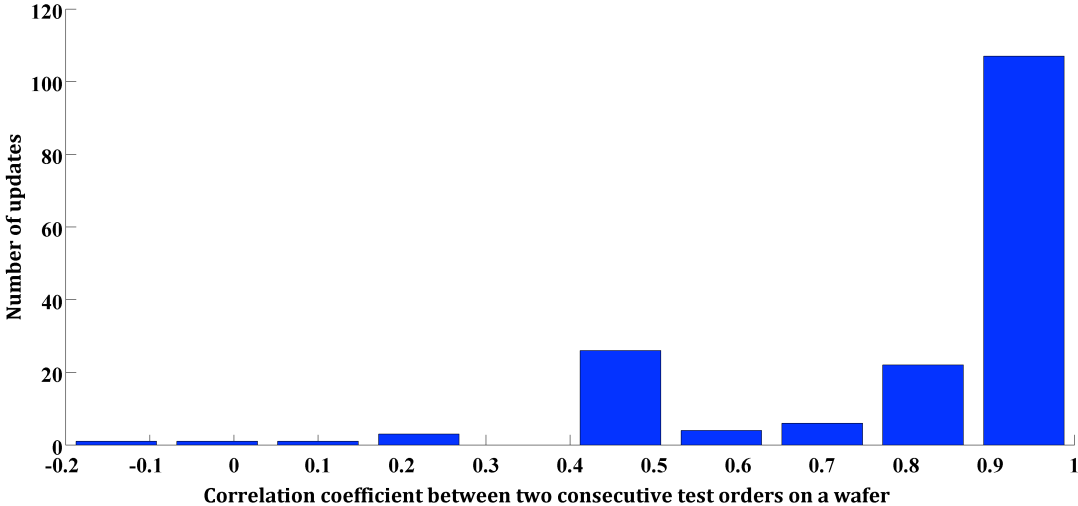


Figure 4.11: Correlation between consecutive test orders on a wafer after application of ATS.

Eliminating tests or patterns reduces the test times for the passing and failing die both, but increases the risk of test escapes. Quality of a product is measured in terms of DPPM and is directly affected if truncated tests or patterns have unique fails. There is always a trade-off between desired quality level and expected TTR which determines the number of patterns or tests which should be considered for elimination. There are two choices to be made when eliminating tests using ATS as described earlier. The choice of when to start elimination can be made after performing an impact of shock analysis for the product data parameters, simulated in the Monte Carlo simulator as described in Section 4.1.3. The choice of number of tests to be eliminated or truncated depends on the DPPM and final TTR goals for the product. This tradeoff between TTR and defect level for Product 1 and Product 2 is presented in Figure 4.12 and 4.13.

Figures 4.12 and 4.13 show the relation between *Relative Defect Level* and *Test Time Reduction* obtained for various numbers of tests and patterns eliminated using truncation ratios ranging from 0 to 1. Relative defect level is chosen to represent product DPPM and is obtained by normalizing the actual DPPM by the maximum number of fails possible for all wafers of that product. For a truncation ratio of 1, i.e. no testing at all, test time reduction is 100% and relative defect level is maximum, i.e. 1. Similarly for truncation ratio of 0 there are no test escapes since the entire test suite is used for testing and TTR is same as that obtained by reordering the tests. As the truncation ratio goes from 0 to 1 test time decreases and the relative defect level increases.

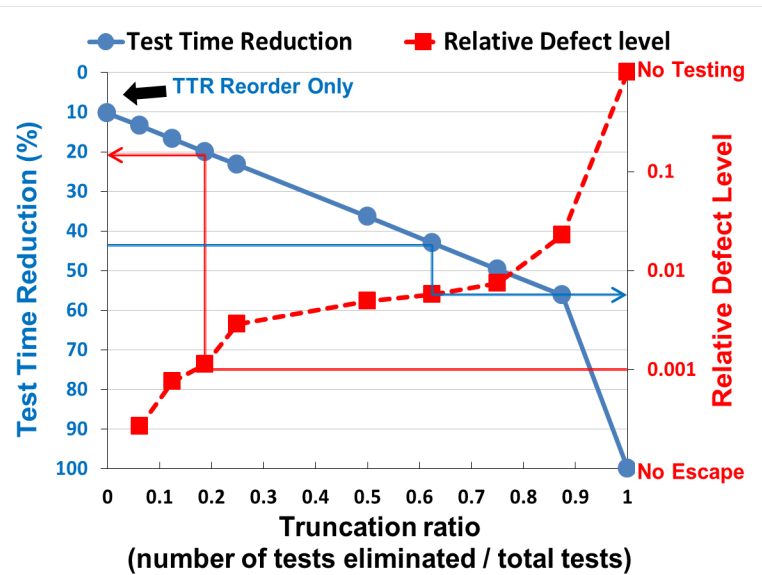


Figure 4.12: Results of application of ATS for Product 1: Tradeoff between TTR and Defect Level.

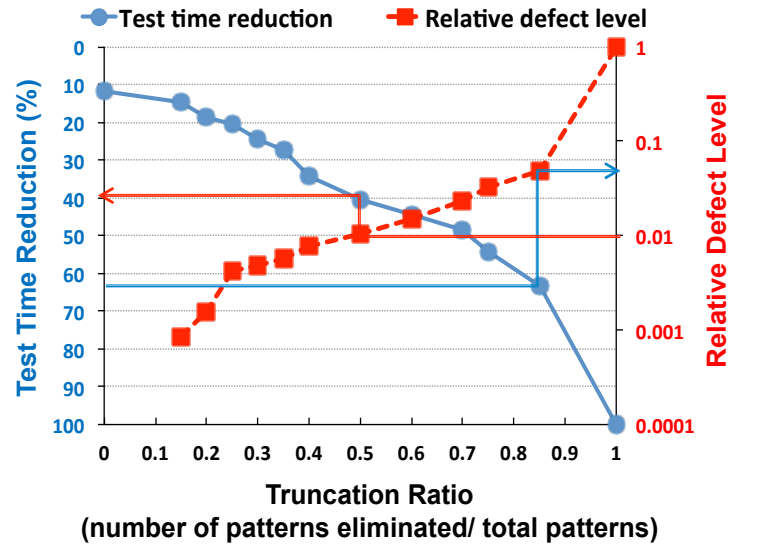


Figure 4.13: Results of application of ATS for Product 2: Tradeoff between TTR and Defect Level.



Figure 4.12 shows the results of application ATS with reordering and elimination for Product 1. A truncation ratio of 0.2 for tests considered for Product 1 gives a test time reduction of about 20% and relative defect level less than 0.001 which corresponds to a DPPM value conforming to the industry requirements for that product. Figure 4.13 shows the results of application ATS with reordering and elimination for Product 2. A truncation ratio of 0.5 for transition fault patterns appearing late in the test application order results in 40% reduction in total test time with a relative defect level less than 0.01 for transition fault patterns conforming with the industry requirements for that product. These TTR values and the relative defect levels are shown by red arrows. A similar example is shown using blue arrows. Both the examples demonstrate the simple linear dependence which could be modeled by Monte Carlo methods outlined earlier.

## Chapter 5

### Conclusions

#### 5.1 Contributions of this work

This thesis presents for the first time an on-tester, adaptive test scheme with dynamic, real-time control of the test flow and test content. This thesis also suggests a new classification for adaptive test methods and takes a step towards continuous adaptation by dynamically reordering and eliminating tests on a per die basis. The adaptive test scheme presented in this thesis is based on the key idea of learning and updating each tests fail rate on a per die basis. The ATS uses Bayesian statistics to model the per test fail rates and responds at test to process variation. The dynamic learning process allows ATS to adapt to screen defects and variation by reordering the tests on the fly. The Bayesian formulation employs simple calculations to update the fail rates which can be done in parallel with the testing process.

The potential benefits of ATS were demonstrated using synthetic data generated by Monte Carlo and real production wafer sort data for products manufactured in semiconductor industry. Reordering tests show an average 56% reduction in test-time-to-fail for Product 1 test data and an average 33% reduction in test-time-to-fail for Product 2. ATS tracks and responds to die level changes of each tests fail rate. The close tracking reduces the test-time-fail variance and allows per-wafer elimination of tests or patterns. Reduction in test-time-to-fail variance reduces the uncertainty in the number of tests required to detect the same failing

die allowing test elimination. Per-wafer elimination of patterns decreases both test-time-to-fail and test-time-to-pass. Product data results show a 20% reduction in total test-time for Product 1 and 40% reduction in test time for Product 2 . Test elimination provides the maximum TTR at the risk of increased DPPM. ATS makes use of an inherent, natural, quality monitor by resetting the test content at the start of wafer and employing per-wafer test elimination. The risks can be contained by the use of the natural monitor.

## 5.2 ATS Production Requirements

The success of adaptive test depends on the details of the ATE platform's support of test pattern loading and test pattern execution, and on the design of the test program itself. For any ASIC test program pattern load is time-consuming and the number of pattern loads must be limited. For reordering different tests/test groups, modifications to ATE architectures are required to avoid test reorder overhead. Test operations like data logging and binning also need to change to incorporate adaptive test flows. For the transition fault patterns considered in the paper, pattern load was a one-time event.

ATS made run-time decisions about the changing the test order using only the already-loaded and standalone pattern groups and hence reordering should not have any overhead. Today, most ATE platforms directly support the STIL programming language in the IEEE 1450 standard [23,24]. STIL modularity allows patterns to be grouped without conflict as described by Taylor and Matson in 1996 [23]. The benefits described in this thesis for die-level adaptive reordering highlight the value of modular STIL programming.

### 5.3 Recommendations and Implications

ATS shows the value of learning from test data on a die-level to the field of adaptive test and shows the importance of data-driven analysis. In light of increasing process variation and defects, ATS shows the benefit of increased frequency of updates to the test order. ATS shows that having one fixed test order to characterize test data and load the test program will result in extra test time and test cost because the test data is actually characterized by multiple test orders for multiple wafers. This work provides a way to determine the frequency of test order updates for future products.

Use of Bayesian statistics enables adaptive data-driven analysis and update of test flow instead of a static analysis afforded by traditional frequentist statistical modeling. Extension of the Bayesian techniques to different prior-posterior pairs, multivariate Bayesian statistics etc. can also prove useful for research in this field.

The Monte Carlo simulator developed for evaluating the performance of ATS and understanding the impact of choice of different seed values represents a robust method that can be used to verify selection of initial parameters in any adaptive scheme. ATS performance was first verified by performing Monte Carlo analysis of synthetic data generated from known distributions which provided valuable insights on the consequences of ATS application in different scenarios. ATS was applied to production test data with unknown natural fail rate distributions. As in Monte Carlo the product data results demonstrated the significant TTR benefit. This is a general procedure which can be applied to any new adaptive test method to understand behavior of the method in different scenarios.

ATS also provides an inexpensive natural monitor for excursions by using wafer-level resets for variances of test fail rate and test contents. ATS can also support test sampling by executing tests on a few sample die instead of eliminating them from the test suite. The natural monitor is an example of how adaptive test schemes can simultaneously reduce test time and preserve quality levels.

Finally, ATS and other dynamic adaptive test methods need support from standard ATE software for wide and successful adoption in industry. The IEEE STIL 1450 extensions for test flow descriptions must anticipate the emergence of adaptive test and be sufficiently flexible to support methods such as die-level changes in the test flow demonstrated in this thesis.

## References

- [1] B. Stine, D. Boning, and J. Chung, "Analysis and decomposition of spatial variation in integrated circuit processes and devices," *IEEE Transactions on Semiconductor Manufacturing*, vol. 10, no. 1, pp. 24–41, Feb 1997.
- [2] R. Aitken, "Defect or variation? characterizing standard cell behavior at 90 nm and below," *IEEE Transactions on Semiconductor Manufacturing*, vol. 21, no. 1, pp. 46–54, Feb. 2008.
- [3] International Technology Roadmap for Semiconductors 2009, Test and Test Equipment. [Online]. Available: <http://www.itrs.net/links/2009itrs/home2009.htm>
- [4] M. Chen and A. Orailoglu, "Test cost minimization through adaptive test development," in *Proceedings of IEEE International Conference on Computer Design*, Oct. 2008, pp. 234–239.
- [5] S. Benner and O. Boroffice, "Optimal production test times through adaptive test programming," in *Proceedings of IEEE International Test Conference*, 2001, pp. 908–915.
- [6] R. Madge, B. Benware, R. Turakhia, R. Daasch, C. Schuermyer, and J. Ruffler, "In search of the optimum test set - adaptive test methods for maximum defect coverage and lowest test cost," in *Proceedings of International Test Conference*, Oct. 2004, pp. 203–212.

- [7] K. Butler and J. Saxena, “An empirical study on the effects of test type ordering on overall test efficiency,” in *Proceedings of International Test Conference*, 2000, pp. 408–416.
- [8] K. Gotkhindikar, W. Daasch, K. Butler, J. Carulli, and A. Nahar, “Die-level adaptive test: Real-time test reordering and elimination,” in *Proceedings of IEEE International Test Conference*, Sept. 2011.
- [9] B. Arslan and A. Orailoglu, “Adaptive test optimization through real time learning of test effectiveness,” in *Design, Automation Test in Europe Conference Exhibition (DATE)*, March 2011.
- [10] K. Nakjima, “A new approach to system diagnosis,” in *Nineteenth Annual Allerton Conference on Communication, Control and Computing*, Oct. 1981.
- [11] S. Huss and R. Gyurcsik, “Optimal ordering of analog integrated circuit tests to minimize test time,” in *28th ACM/IEEE Design Automation Conference*, 1991, pp. 494–499.
- [12] K. Butler and J. Carulli, “Modeling test escape rate as a function of multiple coverages,” in *Proceedings of IEEE International Test Conference*, Oct. 2008, pp. 1–9.
- [13] X. Yu, Y.-T. Lin, W.-C. Tam, O. Poku, and R. Blanton, “Controlling dppm through volume diagnosis,” in *27th IEEE VLSI Test Symposium*, May 2009, pp. 134–139.
- [14] F. Ferhani, N. Saxena, E. McCluskey, and P. Nigh, “How many test patterns are useless?” in *26th IEEE VLSI Test Symposium*, May 2008, pp. 23–28.

- [15] E. Marinissen, A. Singh, D. Glotter, M. Esposito, J. Carulli, A. Nahar, K. Butler, D. Appello, and C. Portelli, “Adapting to adaptive testing,” in *Design, Automation Test in Europe Conference Exhibition (DATE)*, March 2010, pp. 556 –561.
- [16] A. Nahar, K. Butler, J. Carulli, and C. Weinberger, “Quality improvement and cost reduction using statistical outlier methods,” in *IEEE International Conference on Computer Design*, Oct. 2009, pp. 64 –69.
- [17] P. Maxwell, “Adaptive test directions,” in *15th IEEE European Test Symposium (ETS)*, May 2010, pp. 12 –16.
- [18] A. Gelman, J. Corlin, H. Stern, and D. Rubin, *Bayesian Data Analysis, Second Edition*. CRC Press, July 2003.
- [19] S. Kaplan, “On a Two-Stage bayesian procedure for determining failure rates from experimental data,” *IEEE Power Engineering Review*, vol. PER-3, no. 1, p. 43, Jan. 1983.
- [20] Ahmad and Shafaghi, “Equipment failure rate updating Bayesian estimation,” *Journal of Hazardous Materials*, vol. 159, no. 1, pp. 87 – 91, 2008.
- [21] M. L. Bushnell and V. D. Agrawal, *Essentials of Electronic Testing for Digital, Memory, and Mixed-Signal VLSI Circuits*. Boston: Kluwer Academic, 2000.
- [22] C. L. Atwood, “Constrained noninformative priors in risk assessment,” *Reliability Engineering and System Safety*, vol. 53, no. 1, pp. 37 – 46, 1996. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0951832096000269>



- [23] T. Taylor and G. Maston, “Standard test interface language (STIL) a new language for patterns and waveforms,” in *Proceedings of International Test Conference*, Oct 1996, pp. 565 –570.
- [24] G. Maston and T. Taylor, “Layering of the STIL extensions,” in *Proceedings of IEEE International Test Conference*, Nov. 2005.