On Efficiently and Reliably Achieving Low Defective Part Levels *

Li-C. Wang
University of Texas at Austin

M. Ray Mercer
Texas A&M University

Thomas W. Williams
Microelectronics Division, IBM

Abstract

How can we guarantee that a testing method will stably and efficiently achieve a very low defective part level? Traditional testing methods rely upon faults to model all defects. As technology advances, this approach becomes increasingly questionable. If only a subset of defects are modeled as faults, then as fault coverage approaches 100%, the tests will be more and more biased in favor of fault detection. Unfortunately, this reduces the testing efficiency for defects and limits the quality level that we can achieve. In this paper, we propose models for the testing process and suggest a solution which we call "unbiased test generation." We define two types of testing bias, and these new metrics can be used to compare and evaluate test generation methods in practice.

1 Introduction

Testing is performed to weed out the defective parts coming out of manufacturing. Since test generation and test application are limited by available resources like memory and time, generating tests for all defects is infeasible. Instead, traditionally a relatively small set of abstract defects, namely *faults*, are constructed and these faults are *targeted* to produce tests. There can be an enormous number of possible defects in a circuit. To do a good job, a given test set should detect most of them. Since faults do not model all possible defects, the test quality with this approach relies on fortuitous detection of the *non-target* defects [[BUTL90] [BUTL91a] [BUTL91b]].

As the quality demands and circuit sizes increase, the effectiveness of testing approaches based upon a single fault model becomes questionable. For instance, [KAPU92] showed that for the most commonly used single stuck-at fault model, the range in defective part levels can spread over several orders of magnitude. Later, [[PARK94] [WMW95a] [WMW95b]] did extensive studies on this issue and demonstrated that a high fault coverage was insufficient to guarantee an equally high testing quality. The weakness of using the single fault model approach for high quality testing reveals that one model may bias the selection of tests in the way that some non-target defects are missed. One possibility to remedy this bias is to use more fault models to generate more tests (for instance, [MAX92]). Another potential solution was proposed in [WMW95a] where a method called *unbiased testing* was introduced. Encouraging results were established there via a *pseudo-unbiased testing* method that guarantees only site observation. In this paper, we will extend the concept of unbiased testing in terms of a new type of testing bias that was not addressed before. We then show that with the new definition, unbiased testing can "efficiently" and "reliably" achieve an arbitrarily low defective part level. In contrast, we demonstrate the mathematical reasons and practical evidences that traditional testing methods based upon faults result in not only limited testing quality but also higher uncertainty.

In general, this paper intends to provide theoretical answers as well as practical evidences to the following questions: 1) *Why is the effectiveness of a test applied earlier much better than a test applied later?* 2) *Why is it that using traditional testing methods, when the fault coverage is high, test application becomes futile?* 3) *Why is it that traditional testing methods result in higher uncertainty in terms of the testing quality and how can this uncertainty be reduced?* 4) *If, given a testing goal of 1 PPM, how can we reliably achieve that with a reasonable cost?* Without a better understanding of the first three questions, we cannot answer the forth. Previously, researchers have observed the phe-

*This work was supported in part by the Semiconductor Research Corporation under contract SRC 94-DP-388, and in part by the Innovative Science and Technology Office of the Strategic Defense Initiative Organization, administered through the Office of Naval Research under contract # N00014-92-J-1723.

0-7803-2991-0/95 $4.00 © 1995 IEEE

nomenon stated in the first three questions during various experiments. For instance, with the assumption of *surrogates* that are modeled defects not targeted by test generation, [[PARK94] [WMW95a] [WMW95b]] demonstrated those phenomenon via extensive experiments. However, little has been said about the causes before. The theoretical framework developed in this paper explains the phenomenon, and lead us to develop a better testing algorithm — unbiased testing, for the answer to the fourth question.

Through out the paper, we will present various experimental results to confirm the theory developed. Since the analysis involves working on the functional space of a circuit and hence the Ordered Binary Decision Diagram [BRAY86] which requires a tremendous amount of memory resource, we only perform the analysis on small benchmark circuits.

2 Background

A defect is a flaw in a circuit. A fault model is a hypothesis of how defects affect the circuit behavior. Given a fault model, a set of faults is derived, called *target* faults. Then, tests are generated on these faults. Usually, target fault coverage is used as an estimator for defect coverage. Defects are categorized into those which can be mapped directly onto modeled faults and others which cannot. We call the former *target defects*, and the later *non-target defects.* Since detecting a particular non-target defect is not ensured, the accuracy of the estimation for defect coverage depends on fortuitous detection of the non-target defects. Note that there can be an enormous number of defects in a circuit. While lacking a clear uniform model to capture all defects, to study the fortuitous detection, in practice we assume *surrogates.* Surrogates are models of different defects from the faults used for test generation. In our study, we used two sets of surrogates: 1). non-feedback AND bridging and 2). transition (gate delay faults). We then *assume that the surrogate coverage is an accurate predictor of the actual defect coverage.* We chose non-feedback AND bridging faults since feedback bridging faults are easier to detect [[MILL88] [MEI74]].

Test generation involves mainly two issues — the selection of tests and the number of tests selected. For selection of tests, we want a test to be able to detect more defects. Without knowing the test spaces for defects, this goal is not easily achieved. For the size of a test set selected, we perform target fault simulation to compute the current fault coverage, and usually test generation stops when a criteria like 99% fault coverage is met.

To estimate final testing quality, we need to predict the defective part level from this fault coverage, which is usually measured as the number of defective parts per million (PPM). For defective part level prediction, Williams and Brown [WILL81] has the model (WB model) $DL = 1 - Y^{(1-DC)}$, where DL is the defective part level, DC defect coverage, and Y yield (Y is assumed 0.5 in this paper). The yield comes from empirical data on the manufacturing process. In practice, fault coverage FC is used in place of the defect coverage since we do not know how to compute DC. In order to accurately estimate a desired defective part level using FC, we need $FC \approx DC$. If FC differs from DC significantly, which is usual when tests are generated using the single stuck-at fault model, then even FC being close to 100% cannot ensure a particular DL because DC may be still much less than 100%.

3 Previous Results

Many researchers have demonstrated by experiments that test effectiveness declines as the fault coverage approaches 100% [[PARK94] [WMW95a] [WMW95b]]. Figures 10 in Appendix show a typical example. In this study, stuck-at faults were assumed to be the target faults for test generation, and non-feedback bridgings were the surrogates. Thousands of test sets were randomly constructed to compute the mean effect and its standard deviation, using Ordered Binary Decision Diagrams [BRAY86] and the *difference propagation* technique [BUTL91a]. Then, the defective part level at a particular fault coverage is plotted by using surrogate coverage as the defect coverage in the WB model. The curve that assumes that fault coverage is the defect coverage is marked as "WB-Model." First, it can be easily observed that the average defective part level does not follow the "WB-Model" curve when the fault coverage is high. This average curve almost turns into a horizontal line as the fault coverage approaches 100%. This means that as the fault coverage keeps increasing, the defect (surrogate) coverage remains the same. In other words, the tests applied under the high fault coverage *detect almost no defects (surrogates)* and become futile with respect to the improvement of testing quality. To explain the reasons behind those phenomenon, [WMW95a] presented a simple probabilistic model.

3.1 A Simple Model for Testing Process

Assume a fault universe of N_f faults and a defect universe of N_d defects. For a given test, on average

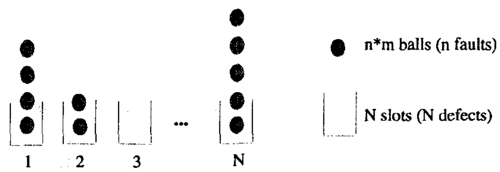

Figure 1: Illustration of the Group Allocation Problem

m_f faults are detected and m_d defects can be detected. Then, the testing process can be modeled by the **Group Allocation Problem**. The problem asks for the number of empty slots at the end of dropping n groups of m balls, group by group, into N slots such that on each dropping, each slot has an equal probability of receiving a ball and no two balls go to the same slot. Figure 1 illustrates the setup of the problem. The classical scheme of this problem for $m = 1$ was treated in [FELL66] and for $m \geq 1$, the average number of empty slots can be easily shown to be $N(1-\frac{m}{N})^n$ which can be approximated by $e^{-\frac{m \cdot n}{N}}$ when n, N are large enough and $\frac{n}{N}$ is fixed. Therefore, if T_l is the test length (the number of tests) applied, then at the end we will have defect coverage $1 - e^{-\frac{m_d T_l}{N_d}}$. When using no fault dropping and $T_l = N_f$, the defect coverage becomes

$$1 - e^{-\frac{m_d N_f}{N_d}} = 1 - e^{\frac{m_d}{\ell}} \text{ where } N_f \cdot \ell = N_d \quad (1)$$

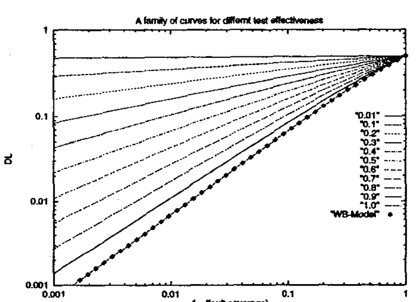

Figure 2: The defective part level curves for various test effectiveness ε

Based upon the model, the *test effectiveness* is defined as $\varepsilon = \frac{m_d N_f}{m_f N_d}$. Figure 2 shows the defective part level curves for ε ranging from 0.01 to 1 with $N_f = 500, m_f = 2, N_d = 5000$. Again, the curve "WB-model" refer to the case when fault coverage is the defect coverage, and it coincides with the curve with $\varepsilon = 1$. These curves divide the area into *low effectiveness regions* (with small ε) and *high effectiveness regions* (with large ε). By comparing Figure 10 in Appendix with this figure, we can observe that the average curve in Figure 10 goes from the high test effectiveness region to the low test effectiveness region as the fault coverage increases. *Unbiased testing* was defined as the case when $\varepsilon = 1$ in [WMW95a].

3.2 Pseudo-Unbiased Testing Method

Since unbiased testing is a theoretical concept, in practice, we search for *pseudo-unbiased* testing algorithms which are an approximation of *unbiased testing*. [WMW95a] proposed one such method and presented encouraging results. In particular, the paper shows that a pseudo-unbiased testing method can result in a lower defect part level and less uncertainty of testing quality for various experiments. To understand the essence of the method, consider test generation for a given fault using traditional methods. For an input vector to be a test, it has to satisfy two requirements — to produce a difference at the fault site and to ensure propagation of that difference to one of the primary outputs. The two requirements correspond to the *fault excitation* and *fault observation* conditions. For different types of defects, the excitation condition can be different. However, if a defect occurs only within the given site, then the observation condition is the same regardless of what type the defect is. Hence, to reduce the bias of test generation based upon a fault model, the method ensure *only* the observation criteria and leave the excitation criteria to be random. In this way, the pseudo-unbiased testing method generates a set of vectors such that the test set *guarantees* observation for every site at least once. We note that the fault model is still involved when computing the fault coverage, but with the new method, faults are not targeted for generating tests any more.

In this paper, we will provide further analysis regarding the problems with traditional testing methods, explain in details the causes, and improve the proposed solution — unbiased testing.

4 The Uncertainty of Testing Quality

[WMW95a] formally studied the relationship between fault coverage and defect coverage under the condition that a fixed number of tests is given. In this section, several new results will be presented using the same model, the Group Allocation Problem. First of all, we should consider not only the case for a given test length as that assumed in [WMW95a] but also the case for a given fault coverage. The difference is that the test length to achieve a fixed fault coverage is no longer a constant but a random variable. Besides, for both cases, we will study their average behavior, as well as their variation in terms of the testing quality. The results in this section provide much more insight about the definition of *test effectiveness* and serve as the basis to understand the remaining sections. The superiority of unbiased testing in terms of reducing the uncertainty of testing quality will be formally proved. This result was suggested only by experiments in [WMW95a]. Due to limited space, we only explain each theorem without proof. Most of the proofs can be derived from the materials in [[FELL66] [HOLS71] [HOLS77] [PARK81]].

Again, let N_f, N_d denote the numbers of faults and defects, and m_f, m_d be the average number of faults/ defects detected by a test, respectively. Let ε be the test effectiveness defined as $\frac{N_f m_d}{N_d m_f}$.

4.1 When the test length is a constant

Equation (1) in previous section stated the expected defect coverage when applying a fixed number (N_f) of tests. The following theorem gives the relationship between the variation of defect coverage DC and test effectiveness.

Theorem 1 *After applying N_f tests, the variance of the number of defects left undetected $N_d(1-DC)$, denoted as $VAR[N_d(1-DC)]$ is $N_d e^{-\varepsilon m_f}[1-\frac{1+\varepsilon m_f}{e^{\varepsilon m_f}}]$ as N_f, N_d become large and $\frac{N_f}{N_d}$ remains a constant.*

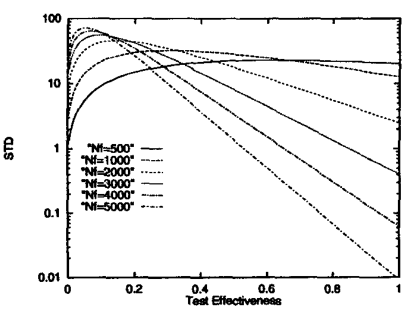

Figure 3: The $Std[N_d(1-DC)]$ vs. ε for various N_f

To better understand the theorem, consider fixed $\frac{m_f}{N_f} = 0.004$ and $N_d = 10N_f$, and ε varies from 0.01 to 1.0. We try to study the change of $VAR[N_d(1-DC)]$ in terms of the change of ε. Figure 3 shows six curves corresponding to the cases $N_f = 500, 1000, 2000, 3000, 4000$ and 5000. Note that the curves are drawn for $Std[N_d(1-DC)] = \sqrt{VAR[N_d(1-DC)]}$ instead of the variance.

For larger N_f, we can see that the case $\varepsilon = 1$ always has less uncertainty, and as N_f increases, this advantage becomes stronger. If we compare Figure 10 with Figure 2, we observe that as fault coverage approaches 100%, the test effectiveness roughly falls into the region with ε around 0.6. Unfortunately, this gives us a relatively larger variance of the defect coverage when N_f is around 500 and hence a higher uncertainty of testing quality for C432.

4.2 When the test length is a random variable

Suppose now the test length T_l is unknown. Instead, we have the goal of achieving a particular fault coverage. Then, T_l becomes a random variable. *What is the new formula for the mean and the variance now?* To answer this question, we need to find the distribution of T_l first. Given $0 \leq FC < 1$, the distribution of T_l to achieve at least FC fault coverage is normal with mean and variance as the following ($O(1)$ means a small constant).

$$EXP(T_l) = \frac{N_f}{m_f} \ln \frac{1}{1-FC} + O(1) \qquad (2)$$

$$VAR(T_l) = \frac{N_f}{m_f^2}\left(\frac{FC}{1-FC} - \ln \frac{1}{1-FC}\right) + O(1) \qquad (3)$$

The proof for $m_f = 1$ can be found in [HOLS71]. The extension to $m_f > 1$ can then be obtained by considering each of m_f balls in a group independently. With the help of above results, we can derive the following theorem.

Theorem 2 *Let ε be defined as usual. Let T_l be the test length applied, which has a distribution described above. As N_f, N_d become very large, and $\frac{N_d}{N_f} = \ell$ remains a constant, asymptotically we have*

$$EXP(1-DC) = (1-FC)^{\varepsilon} \qquad (4)$$

$$\begin{aligned} VAR[N_d(1-DC)] = N_d EXP(1-DC) \times \\ \{1-[\ell(\varepsilon m_f)^2+\varepsilon](1-FC)^{\varepsilon}\ln(\tfrac{1}{1-FC}) \\ +\ell(\varepsilon m_f)^2(\tfrac{FC}{1-FC})\} \end{aligned} \qquad (5)$$

The proof for Theorem 2 is a little bit complicate. Here we omit the details. Equation (4) although is different from equation (1) in Section 3.1, they both define the same set of test effectiveness curves as shown in Figure 2. Therefore, the definition of *test effectiveness* has the same meaning for both cases and hence its definition is independent of whether the test length is a constant or a random variable. We note that fault targeting is not addressed here. With fault targeting, a test is specifically ensured to defect a particular fault and hence the assumption of average detection of m_f faults is not appropriate. However, this problem can be solved by partitioning all faults into those explicitly being targetd and those fortuitously detected. [WMW95a] discussed the analysis. Since fault targeting does not affect the conclusions drawn in this paper, we will omit its discussion.

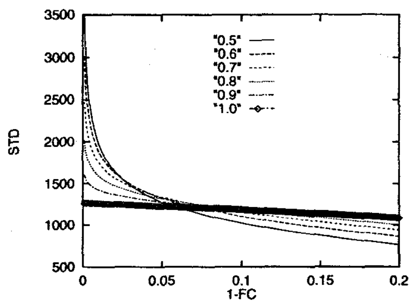

Figure 4: STandard Deviation curves $STD[N_d(1 - DC)]$ for various ε for $FC \geq 80\%$

To get a better insight of equation (5), we restrict our attention to the situation when $\varepsilon \geq 0.5$. The reason is that although low test effectiveness may result in a smaller variance, it is less interesting to us due to its weakness of reducing the defective part level. Figure 4 presents the curves of $STD[N_d(1 - DC)]$ as before for various $\varepsilon \geq 0.5$ and $FC \geq 80\%$, given that $N_f = 1000, m_f = 4, N_d = 10000$. It is interesting to note that the case $\varepsilon = 1.0$ has the highest uncertainty before 95% fault coverage but becomes relatively much smaller than the others when the fault coverage approaches 100%.

The most interesting conclusion observed from Figures 3 and 4 is that ***by increasing test effectiveness ε, the uncertainty of testing quality can be reduced as the fault coverage approaches 100%.*** This result was partially demonstrated by the experiments in [WMW95a]. We emphasize that this superiority is consistent for both the case of fixed test length and the case of fixed fault coverage (providing that the fault coverage is high), and high fault coverage is the domain of interest for good quality levels.

4.3 Experiments

Figures 5 and 6 present the results obtained through experiments on circuits C432 and C499 using transition surrogates (From our experience, transition surrogate set gives larger variance than bridging [WMW95a]). As before, we construct thousands of test sets to compute the average defective part level and its standard deviation at a given fault coverage. We compare the results for two test generation methods, the traditional testing using the stuck-at fault model and the pseudo-unbiased testing addressed in Section 3.2. It is clear from the figures that the pseudo-unbiased testing method can produce more certain testing quality than the stuck-at fault approach, and this difference becomes larger as the fault coverage increases. We note that the pseudo-unbiased testing method was originally designed to enhance test effectiveness, and as a result the uncertainty was also reduced. We will discuss this aspect in more details below.

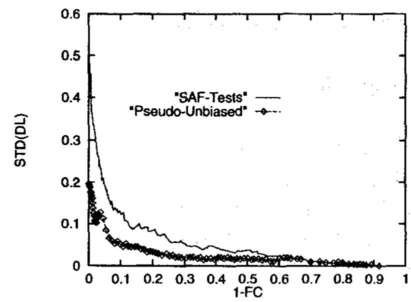

Figure 5: STandard Deviation curves $STD(DL)$ vs. fault coverage $1 - FC$,transition surrogates,C432

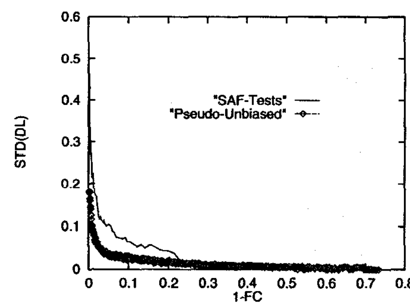

Figure 6: STandard Deviation curves $STD(DL)$ vs. fault coverage $1 - FC$,transition surrogates,C499

5 Bias, Efficiency, Unbiased Testing

Our purpose in this section is to explore the meaning of "unbiased testing" further, and formally unveil the concept of *testing bias*. By carefully distinguish two types of testing bias (or bias for simplicity), we obtain more insight about the testing process. Then, we propose a new concept, *test efficiency*, that is similar to test effectiveness. Those new concepts helps to understand more about the problems with respect to traditional testing methods and lead to define a better testing algorithm later.

5.1 Test Generation Bias (T-Bias) vs. Fault Simulation Bias (F-Bias)

Recall that test effectiveness is defined as $\varepsilon = \frac{m_d N_f}{m_f N_d}$. From the test effectiveness curves in Figure 2, we see that the case $\varepsilon = 1$ is the one where the "average ability" for a test to reduce the fault coverage is the same as its "average ability" to reduce the defect coverage. In contrast, "bias" can be considered as *the "difference" between these two average abilities.* Here, we explicitly define this concept.

Definition 1 *Define the Fault Simulation Bias (F-bias), denoted as β_F, to be $1-\varepsilon$, where ε is the test effectiveness.*

Since $\varepsilon \geq 0$, $\beta_F \leq 1$. Usually, we don't have $\varepsilon > 1$ since that implies "defects are easier to detect than faults," and if $\varepsilon > 1$, most of the defects will be detected before the fault coverage reaches 100%. However, when we assume surrogates for experiments, $\varepsilon > 1$ can happen simply because surrogates does not capture all possible defects. As long as β_F is not positive, a high fault coverage can guarantee a high testing quality. Therefore, a test set with $\beta_F = -0.1$ and a test set with $\beta_F = 0$ both are good test sets. However, since $\beta_F = -0.1$ means more favor of defect detection, we may generate more tests than required for the desired quality level if fault coverage is used as the predictor for quality.

Now, let's again compare the result in Figure 10 with Figure 2. We see that ε for that particular setting is not a constant as the fault coverage changes, but a variable that decreases as the fault coverage increases. This observation indicates that another type of "bias" is associated with the test set, which is not modeled by the definition of β_F (or equivalently not modeled by the concept of test effectiveness). This bias reflects *the difference between the two average abilities to reduce defect coverage at two different fault coverages.* To better understand this new type of bias, we consider a partition of the defect space with respect to the test effectiveness. Let D be the defect space partitioned into k disjoint subset $D_1 \ldots D_k$. Hence, if the cardinality $|D| = N_d$ and $|D_i| = N_i$ for $1 \leq i \leq k$, then $N_d = N_1 + \cdots + N_k$. We also denote the test effectiveness for D_i as ε_i, and assume that $\varepsilon_1 > \cdots > \varepsilon_k$. Note that there is no equality when comparing ε_i with ε_{i+1} since if they are equal, D_{i+1} should be merged into D_i. Under the setup, we define the new type of bias as the following.

Definition 2 *Let $p_i = \frac{N_i}{N_d}$. Let $\varepsilon_{ave} = \Sigma_{i=1}^{k} p_i \varepsilon_i$. Then, we define the Test Generation Bias (T-bias) to be $\beta_T = 1 - \frac{\varepsilon_k}{\varepsilon_{ave}}$ where $\varepsilon_1 > \varepsilon_2 > \cdots > \varepsilon_k$.*

The meaning of T-bias is simple. β_T measures the the ratio between the minimum test effectiveness and the average test effectiveness over the defect subsets. Since the minimum is less than or equal to this average, β_T is always a number between 0 and 1, and is 0 when the equality holds. When the minimum is the same as the average over defect subsets, the detection ability over all defects is uniform, i.e no T-bias is present. In this case, the overall test effectiveness will be a constant as the fault coverage increases. In general, we want β_T as close to zero as possible.

With the definition of F-bias and T-bias, we extend the notion of unbiased testing below.

Definition 3 *A testing method is called* **F-unbiased** *if its resulting $\beta_F = 0$ and is* **T-unbiased** *if $\beta_T = 0$. An* **Unbiased Testing** *method is the one that is both F-unbiased and T-unbiased.*

So far, we have not explained the reason why we used the name "Test Generation Bias" and "Fault Simulation Bias." To understand the essence of T-bias and F-bias, we demonstrate a few experiments on the benchmark circuits C432 and C499. Via realistic examples, we will illustrate that F-bias is indeed changeable by *changing the way of computing the fault coverage* and T-bias can be reduced by an *improved test pattern generation algorithm.*

5.2 Experiments on β_F and β_T

p_i	0.91	0.07	0.02
ε_i	1.216	0.502	0.0

Table 1: Bridging surrogate set partitioning using SAF test sets, C432

In the previous section, we define different types of bias and extend the concept of unbiased testing further. But how well does the model match to reality? Does bias really occur? Can we improve the testing

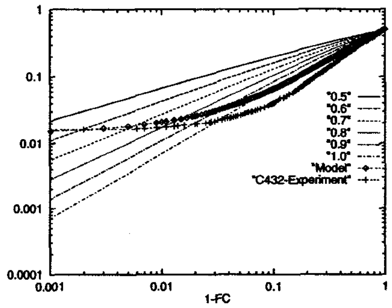

Figure 7: Results on C432 using bridging surrogates and SAF test sets

quality by reducing bias in practice? To answer these questions, we need experiments. Let's first check the accuracy of the model.

There are four numbers we need to know in order to compute the test effectiveness ε_i for subset i. They are N_f, m_f, N_d, m_d. Given a subset of surrogates (or faults), we apply a number r of different test sets generated using the same method (either traditional stuck-at fault method or pseudo-unbiased method), each with v test vectors. Then, for surrogates $1, \ldots, n$, we collect for each surrogate, the number of times it is detected and get the sequence of numbers $s_1, \ldots, s_n$, respectively. We then compute the average number of times for a surrogate to be detected by a test as $(\Sigma_{i=1}^{n} s_i)/(r \cdot v \cdot n)$. In this way, we compute m_f, m_d (m_d now is for the surrogates because we assume that surrogates are the defects). N_f, N_d are just the size of the fault set and the size of the surrogate set. To partition the surrogate set into subsets, we sort the sequence of numbers $s_1, \ldots, s_n$ and then group those surrogates in a way that two surrogates stay in the same subset if their numbers of times being detected are close.

As an example, let's consider circuit C432 with bridging surrogates. The test sets are generated based upon the stuck-at faults. Table 1 shows the resulting partition and its corresponding test effectiveness. There are three subsets whose sizes are normalized to $0.91, 0.07$ and 0.02. As we can see in this example, most of the bridging surrogates are very easily detected, but there are 2% of them with test effectiveness of zero, i.e. never being detected. Figure 7 then presents a comparison between what the model gives and what we obtained through experiments. It can be seen that the model, although not exactly matching the actual curve, captures what happens in practice. The figure also contains curves with ε ranging from 0.5 to 1.0. We note that similar results were obtained for

	F-bias β_F		T-bias β_T	
	SAF Tests	Unb Tests	SAF Tests	Unb Tests
BSs	−0.14	−0.22	1.0	1.0
TSs	0.09	−0.04	0.872	0.782

Table 2: Comparison of the two methods in terms of bias, on C432 (BS=bridging surrogate, TS=transition surrogate)

	F-bias β_F		T-bias β_T	
	SAF Tests	Unb Tests	SAF Tests	Unb Tests
BSs	−0.0089	−0.119	1.0	1.0
TSs	0.366	0.234	0.871	0.709

Table 3: Comparison of the two methods in terms of bias, on C499

C432 with transitions and for C499 with both bridgings and transitions.

We now consider the ability of reducing the bias for a given testing method. We compare the resulting F-bias and T-bias for the single stuck-at fault method and the pseudo-unbiased method described in Section 3.2. We note that the pseudo-unbiased method was originally designed to enhance test effectiveness. Table 2 and Table 3 present the results for C432 and C499, respectively. For all four cases (each case involve two methods and two types of bias), we verify that the pseudo-unbiased method indeed reduces the testing bias from that given by the stuck-at fault method. For F-bias, the pseudo-unbiased method has a smaller bias value in all four cases, and for T-bias, although there is no difference with respective to the bridging surrogates, the pseudo-unbiased method reduces T-bias for the transition surrogates. Since *less F-bias implies higher test effectiveness* and *less T-bias implies stronger ability to maintain that test effectiveness*, we expect that the pseudo-unbiased method should give us better results for all four cases. This is indeed true. Figures 11 and 12 ("Unb-" denotes the pseudo-unbiased method) in Appendix show the comparison of the average defective part level curves for these four cases.

We will conclude this section by giving the reason for using the name "Fault Simulation" and "Test Generation" bias. We first note that F-bias can be reduced (or even removed) by using a different fault simulation strategy. As an example, let's consider circuit C499 with transitions and the pseudo-unbiased test

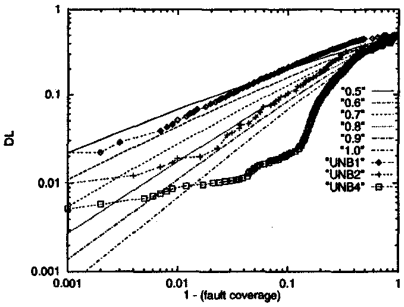

Figure 8: Changing ε and F-bias by changing the way of computing the fault coverage, C499 with transitions

sets. We perform experiments on three different ways of computing the fault coverage. First, as usual we consider the number of faults detected *at least once* being divided by the total number of fault as the fault coverage. Additionally, we consider *detected at least twice* and *detected at least four times* as the ways to compute the fault coverage, respectively. Figure 8 show the resulting comparison. The curve marked by "UNBi" represented the case where a fault is considered "detected" if it is detected at least i times. We see that as i becomes larger, the resulting defective part level curve falls into higher test effectiveness regions (smaller F-bias). Also, we see that the decline of test effectiveness does not change much in all three cases. Hence, *the way of computing the fault coverage can reduce F-bias but cannot remove the decline of test effectiveness, i.e T-bias.* In the next section, we will prove that as the fault coverage approaches 100%, the overall test effectiveness will be dominated by the smallest test effectiveness implied by the partition of the defect set. Since this "non-uniformity" of test effectiveness depends on test generation method itself (as shown in Table 3), we give it the name "Test Generation Bias."

5.3 Test Efficiency When a Fixed Fault Coverage is Given

Test effectiveness characterizes the average ability to reduce the defect coverage as the fault coverage increases. In this section, we define the concept of the "instantaneous ability" to reduce the defect coverage at a particular fault coverage. We call it **Efficiency** to distinguish from test effectiveness. Also, in this section, we assume that a fixed fault coverage is the goal of testing.

Definition 4 *We define the* **test efficiency** *with respect to the change of fault coverage to be* $\xi_{DL} = \frac{\mathrm{d}\ln(DL)}{\mathrm{d}\ln(1-FC)}$.

The follow lemmas connect ξ_{DL} to ε assuming that no T-bias is presented, i.e $\beta_T = 0$.

Lemma 1 *As fault coverage FC approaches 100%,* $\xi_{DL} = \frac{\mathrm{d}(\ln DL)}{\mathrm{d}\ln(1-FC)} \longrightarrow \varepsilon$.

Proof. From equation (4) in Theorem 2 $\ln(DL) = \ln[1 - Y^{(1-FC)^{\varepsilon}}]$. The remaining steps involve applications of the L'Hospital's Rule. □

Lemma 2 *ξ_{DL} is always less than or equal to ε, and it increases as the fault coverage FC increases.*

Proof. Let $h = (1 - FC)^{\varepsilon}$. Then, again we will have $\xi_{DL} = (\frac{-1}{1-Y^h} \cdot e^Y \ln h \cdot h)\varepsilon = (X)h$ (for $h = (1 - FC)^{\varepsilon}$). If we let $h \rightarrow 1$, then $\xi \rightarrow 0$ since $\ln h \rightarrow 0$. We know from lemma (1) that as $h \rightarrow 0$, $\xi_{DL} \rightarrow \varepsilon$, and hence $X \rightarrow 1$. The only thing needs to be shown is that X is monotonically increasing as h decreases. We can prove this by showing that $\frac{\mathrm{d}X}{\mathrm{d}h} < 0$. □

The above illustrates the practical meaning for the test effectiveness ε. Lemmas 1 and 2 states that ε is the upper bound on the order of magnitude reduced on the defective part level when increasing the fault coverage by one order, and in the limit, it is equal to the test effectiveness ε. The practical meaning for this is that *if we increase the fault coverage by one order, we cannot do better than that amount with respective to the reduction of defective part level.* However, if the fault coverage is high enough, the order reduced on the defective part level is roughly ε.

In the following, we consider the case when $\beta_T > 0$.

Lemma 3 *Let the defect space be partitioned into k disjoint sets with cardinalities $N_1 + N_2 + \cdots + N_k = N$ and their test effectiveness be ordered as $\varepsilon_1 > \varepsilon_2 > \cdots > \varepsilon_k = \varepsilon_{min}$. Then, as the fault coverage FC approaches 100%,* $\xi_{DL} = \frac{\mathrm{d}\ln(DL)}{\mathrm{d}\ln(1-FC)} \longrightarrow \varepsilon_{min}$.

Proof. Let $p_i = \frac{N_i}{N_d}$ for $i = 1 \ldots k$. Then, $DL = 1 - Y^{[\Sigma_i(p_i)(1-FC)^{(\varepsilon_i)}]}$. The remaining steps involve applications of the L'Hospital's Rule. □

In general, if there is a subset of defects with dominating size relative to others, then most of the testing process will be dominated by the test effectiveness of that subset (this can be easily proved). However, *when the fault coverage is high enough, the one with the minimum test effectiveness emerges to become the dominating subset.* As an example, consider the partition in Table 1 for C432. Figure 9 shows the decline (eventually go to zero) of ξ_{DL} as the fault coverage approaches 100% for this case from experimental

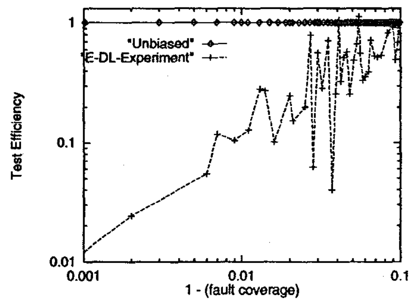

Figure 9: ξ_{DL} vs. FC, C432, bridgings

data. In contrast, the curve for the unbiased testing is also drawn. The undesired drop of ξ_{DL} indicates that the zero test effectiveness subset becomes more active as FC approaches 100% (as predicted in Lemma 3), which leads to the result shown in Figure 7.

In summary, test efficiency measuring how well testing performs at a given fault coverage. It depends on F-bias in average and on T-bias as the fault coverage close to 100%. Therefore, to improve test efficiency, we ought to reduce both types of bias. We conclude this section with the following two theorems.

Theorem 3 *Suppose a testing method M generates a test set T such that for most of the defects, the test effectiveness is ε_1 but for some of them is ε_2, and $\varepsilon_1 \gg \varepsilon_2$. Then, in general, we can perform our testing no better than ε_1 and as the fault coverage approaches 100%, we can perform no better than ε_2.*

Theorem 4 *The ability to detect defects for an unbiased test set ($\beta_T = 0$ and $\beta_F = 0$) is a constant, and independent of the fault coverage.*

6 Efficiently Achieve an Arbitrarily Low DL

Suppose that we are given the testing goal 1 PPM. How are we going to achieve that goal? Suppose we have a testing method in mind. Then, the problem can be divided into two parts: *1). Can the method achieve 1 PPM, providing no resource limitation? 2). If yes, can we afford the cost by using the method to achieve 1 PPM?* First let us examine the potential of traditional testing methods with respect to the first question. Traditionally, tests are generated based upon faults. Since faults do not model defects perfectly, some defects are missed. For those defects not captured by any faults, their test effectiveness is obviously zero. Therefore, in Lemma 3 above, $\varepsilon_{min} = 0$. As the fault coverage approaches 100%, the defective part level reduction provided by more tests is zero. As a result, *the testing quality that can possibly be achieved by this method is limited.*

Let us think about the above two questions in terms of the concept "bias." We see that the T-Bias β_T is very related to the first question, and the F-Bias β_F the second question. If a method has $\beta_T = 0$, then at least we know that the ability to reduce the defective part level is not dropping while the fault coverage increases. Hence, as long as we observe that the test effectiveness is not zero (at least some defects are detected by the first few tests), we can proceed until 1 PPM is achieved. On the other hand, if the method also has $\beta_F = 0$, then we know that *the decreasing rate of the defective part level is as large as the increasing rate of the fault coverage* as the fault coverage approaches 100%. Therefore, we not only can get there but also get there *as fast as we can on the fault domain.* This idea has been captured implicitly in the new definition of unbiased testing before (see Definition 3) and explicitly proved in the previous section. In summary, with the proof that unbiased testing results in less quality uncertainty in Section 4, we conclude our main theorem below.

Main Theorem: *Unbiased testing can achieve an arbitrarily low defective part level efficiently and reliably.*

7 Conclusion

Test Generation Bias and Fault Simulation Bias are two causes that limit the testing quality. A good test generation algorithm should produce test set with as less bias as possible. When there is neither T-bias nor F-bias, a testing method is unbiased, and an unbiased method can efficiently and reliably achieve desired quality. In practice, we design ATPG algorithms to approximate unbiased testing, and such an approximation is called pseudo-unbiased testing. A pseudo-unbiased method with good F-bias reduction was proposed in [WMW95a]. An enhancement to further reduce T-bias is currently under development.

Acknowledgment: The authors like to thank Professor Martin D.F. Wong at the Computer Science Department, University of Texas at Austin, for providing valuable material on the Group Allocation Problem.

References

[BRAY86] R.E. Bryant, *Graph-based algorithms for Boolean function manipulation*, IEEE Trans. on Computers, vol C-35, no. 8, Aug. 1986, pp. 677-692.

[BUTL90] K. M. Butler and M.R. Mercer, *The Influences of Fault Type and Topology on Fault Model Performance and the Implications to Test and Testable Design*, Proc. 27th ACM/IEEE DAC, Orlando, FL, June 24- 28, 1990, pp. 673-678.

[BUTL91a] K. M. Butler and M. R. Mercer, *Assessing Fault Model and Test Quality*, Kluwer Academic Publishers, 1991, ISBN 0 - 7923 - 9222 - 1.

[BUTL91b] K. M. Butler and M. R. Mercer, *Quantifying Non- Target Defect Detection by Target Fault Test Sets*, Proc. of The European Test Conference, Munich, Germany, April 10- 12, 1991, pp. 91-100.

[FELL66] Feller, W *An Introduction to Probability Theory and Its Applications*, 3rd ed. Wiley, New York 1966.

[HOLS71] Holst, L. *Limit Theorems for Some Occupancy and Sequential Occupancy Problems*, Ann. Math. Statist. 42, 1971, pp. 1671-1680.

[HOLS77] Holst, L. *Some Asymptotic Results for Occupancy Problems*, the annal of prob. Vol. 5 No. 6, 1977, pp. 1028-1035.

[KAPU92] R. Kapur, J. Park, and M. R. Mercer, *All Tests for a Fault are Not Equally Valuable for Defect Detection*, Proc. 1992 International Test Conference, Baltimore, MD, September 20-24, 1992. pp. 762-769.

[PARK94] Jaehong Park, Mark Naivar, Rohit Kapur, M. Ray Mercer and T. W. Williams, *Limitations in Predicting Defect Level Based on Stuck-at Fault Coverage*, the VLSI Test Symposium 1994.

[MAX92] Maxwell, P. C. and Aitken, R. C. *I_{DDQ} Testing as a Component of a Test Suite: The Need for Several Fault Coverage Metrics*, Journal of Electronic Testing 3, 1992, pp. 305-316.

[MEI74] Mei, K.C. *Bridging and stuck-at faults*, IEEE Tran. on Computers, Vol C-23, no. 7, 1974.

[MILL88] Millman, S. D. and McCluskey, E. J. *Detecting bridging faults with stuck-at test sets*, Proc. International Test Conference 1988.

[PARK81] Park, C. J. *On the distribution of the number of unobserved elements when M-samples of size N are drawn from a finite population*, Comm. Statist. – Theory and Method, A10(4), 1981, pp. 371-383.

[WMW95a] Wang, Li-C., Mercer, M. Ray, Williams, T. Thomas *Reducing Defective Part Level Via Unbiased Test Set*, in Proc. European Design and Test Conference, Paris, France, 1995.

[WMW95b] Wang, Li-C., Mercer, M. Ray, Williams, T. Thomas *On the Decline of Testing Efficiency as the Fault Coverage Approaches 100%*, in Proc. VLSI Test Symposium, Princeton, NJ, 1995.

[WILL81] Williams, T. W. and Brown, N. C. *Defect level as a function of fault coverage*, IEEE Trans. on Computers, Vol. C-30, no. 12, 1981.

Appendix:

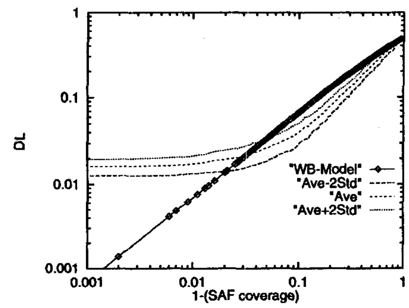

Figure 10: A typical example of defective part levels

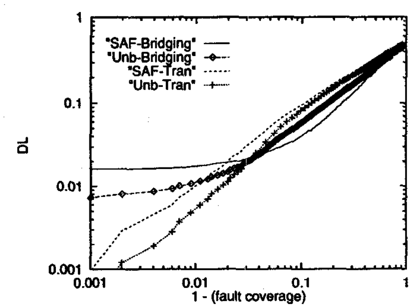

Figure 11: Comparisons of the two methods, C432

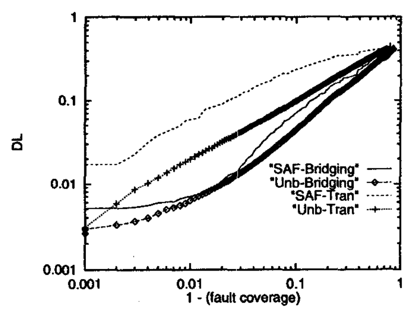

Figure 12: Comparisons of the two methods, C499