Ming Hsieh Department of Electrical Engineering
Viterbi School of Engineering
University of Southern California

Error-rate Estimation with Ones Counting

Zhaoliang Pan and Melvin A. Breuer

Abstract—**With VLSI circuit feature size scaling down, it is becoming more difficult and expensive to achieve a desired level of yield. Error-tolerance, that is, employing defective chips that occasionally produce erroneous yet acceptable results in targeted applications, has been proposed as one way to increase effective yield. In the domain of error-tolerance, defective chips are characterized by certain criteria set by various applications. Error-rate, namely how frequent errors occur at the output, is one such criterion. In this report we focus on the following problem: Given a combinational logic circuit that is defective and hence occasionally produces an erroneous output, how can one determine the error-rate of each output line by using ones counting? This problem was previously considered by others but their analysis has some inaccuracies. This report presents both a more accurate analysis and new estimator of error-rate.**

I. INTRODUCTION

Classical VLSI post-manufacturing tests partition chips into two categories, namely those that pass and those that fail. Presumably, chips that fail have defects and are discarded, while the others are sold to customers. In this report we will use the terms *fail* and *defective chip* interchangeably. The fraction of defect-free chips among all manufactured chips is called the *yield* of a VLSI manufacturing process. With minimal feature sizes scaling down to below 90nm, it is becoming more difficult, hence expensive, to achieve a desired level of yield [1]. Defective chips contribute to a large economic loss to chip manufacturers, and increased prices to consumers.

In an attempt to increase yield, designers employ several techniques such as fault-tolerance, defect-tolerance and design-for-manufacturability [2] [3] [4]. Another somewhat radical and new technique orthogonal to these is called error-tolerance [5] [6]. A chip is said to be *error-tolerant* (ET) with respect to an application or system if 1) it contains defects and hence occasionally produces errors, and 2) using this chip in a given application/system produces acceptable results to the

end user. Because a fraction of defective chips are used in systems rather than being discarded, error-tolerance increases effective yield, and hence potentially profit.

Many examples of systems that can operate in an error-tolerant mode reside in the domain of multi-media and man-machine interfaces. Some simple applications are briefly described below.

While the application of error-tolerance in digital systems probably dates back several decades, one of the earliest documented applications we have encountered appears in a 1995 patent:

> It is herein recognized that in an audio signal processing system, a DRAM chip for storing digitized audio signals and selected to include at least one inoperative memory location, is acceptable for use as a storage medium in that no noticeable error is produced on playback of the recorded signal due to the sampling rate of the audio signal and due to the relatively low rate of defects allowed. Furthermore, the use of a less-than-perfect DRAM chip for storing audio information is acceptable due to the substantially non-critical nature of audio signals, as opposed to the extremely critical nature of computer data. [7]

The following statement appears in another patent dealing with a telephone-answering machine: "The use of 'audio RAMs' (ARAMs) is also known in the answering machine art, with ARAMs being RAMs that are allowed to have a small number of defective bits, in order to allow the use of lower-cost integrated circuit memory chips."[8] Both statements embody the concept of error tolerance, i.e., that some defective circuitry can produce acceptable performance.

An MPEG circuit is another example where error-tolerance is applicable. The effect of defects in some functional blocks of a MPEG encoder, such as the motion estimation block, on the performance of the circuits has been studied [9] [10]. The results indicate that for about 50% of those defects that can be modeled as a single or double stuck-at fault, the resulting defective MPEG circuits generate acceptable performance. Other examples of error-tolerant applications have been reported elsewhere [12] [15]. Error-tolerant chips range from low priced memories to high priced audio and video processing chips. In the near future, many system-on-chip

products will be excellent candidates for error-tolerant applications, such as those containing embedded language translators as well as audio, touch, smell and/or video processors.

Given an analog system, one can inject "errors" into the circuit at some node by applying noise at that node and measure attributes of the output, such as dB loss, MOS value[1], or instability. In a digital circuit, one can introduce errors on an internal signal line and measure the resulting impact on the data at the output bus. Two measures that have been used to quantify output errors are error-rate and error-significance [5] [6]. A formal definition of error-rate is provided in [12]. *Error-rate* indicates how often on average an error is seen at the output when random input patterns are applied. If the circuit is combinational and contains a defect, and the patterns are uniformly selected over the input space, then the error-rate is equivalent to the fraction of the input space that detects the defect, i.e., produces an erroneous output. *Error-significance* indicates the magnitude of an error, when the associated pattern represents some numeric quantity. Depending on the application either one or both of these measures might be applicable.
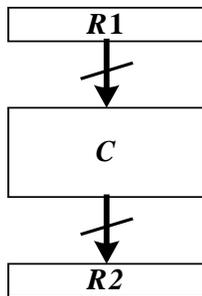
Fig. 1. Simplified BIST architecture

In this report we focus on estimating the error-rate of a defective circuit having static defects. Based on their error-rate, defective chips can be binned into categories. For example, two error-rate thresholds, say r_{th1} and r_{th2}, where $0 < r_{th1} < r_{th2} < 1$, can be used to divide the range of error-rate into three parts, i.e., $(0, r_{th1}]$, $(r_{th1} < r_{th2}]$ and $(r_{th2}, 1]$. Defective chips can now be partitioned into three categories, namely category I with error-rate in the range $(0, r_{th1}]$, category II in $(r_{th1} < r_{th2}]$, and category III in $(r_{th2}, 1]$. Chips in category III (high error-rate) are discarded; chips in category II are sold at a large discount in price; and those in category I are sold at a moderate discount.

Our technique for estimating error-rate requires that we apply a large number of pseudo-random patterns to a targeted circuit. Most complex modern circuits employ either a full-scan DFT or a BIST methodology [16] [17]. While either of these test methodologies can be used to implement our error-rate estimation technique, we will describe our work in the context of a BIST methodology since this simplifies the discussion. A simplified version of a BIST architecture used to test combinational logic is shown in Fig. 1. In normal operation, register $R1$ receives data in parallel and drives the block of logic C. C in turn passes data to $R2$. In the design-for-test process, these registers are modified to operate differently when in the test mode. In the test mode, $R1$ operates as a pseudo-random pattern generator (PRPG), and $R2$ as a multiple input signature analyzer (MISR). The operation of the test process is as follows. First the registers are initialized to some known states. Then the registers are clocked L times, where L is often a very large number. At the end of this process, if the state of the MISR is the same as some pre-computed value, then the circuit C is said to pass (is good or defect free), otherwise it fails. The idea here is that if there is a defect in the circuit, for at least one of the L input patterns the corresponding output pattern would be in error, and once an error enters the MISR, with a very high probability the state of the MISR remains in error even as new data arrives at its inputs. With some assumptions regarding the distribution of output patterns, this technique can be effectively applied to some forms of sequential circuits, such as feed-forward pipelines.

Thus, to implement the binning process, it is important to efficiently bound the error-rate of a defective chip. Note that the test methodology discussed in this report does not employ a fault model, test pattern generation or fault simulation. A technique for estimating the error-rate of a block of logic using signature analysis was first proposed in [11]. Subsequently, Pan and Breuer significantly expanded on the theoretical aspects of this methodology [12]. Shahidi and Gupta then proposed a variant of this work using ones counting rather than signature analysis [13]. Unfortunately, their analysis implied that error-rate estimation via ones counting used significantly more resources than that required via signature analysis. As will be explained later, their results are counter intuitive. In this report, we briefly review the analysis presented in [13], and present a new analysis for this ones counting technique and a new estimator of error-rate. The results from our analysis show that ones counting and signature analysis techniques use comparable resources for obtaining the same degree of accuracy in error-rate estimation.

This report is organized as follows. Section 2 reviews the ones counting error-rate estimation work presented in [13]. In Section 3, we present our new analysis of this technique. A new estimator is proposed and the statistical characteristics of the estimator are presented. Several special cases of using the ones counting technique are described in Section 4. In Section 5, simulation results are presented. The problem of chip classification based on ones counting error-rate estimation is discussed in Section 6. Section 7 presents the experimental results of chip classification, and Section 8 concludes the report.

II. Review of Previous Work

The error-rate estimation techniques proposed in [5], [6],

---

[1] MOS refers to mean opinion score, and indicates the quality of audio speech after processing by some CODEC.

[12] and [13] are illustrated using a BIST architecture, though a scan architecture is equally applicable. The register driving the circuit-under-test (CUT) is a PRPG, which is usually implemented via a linear feedback shift register (LFSR). In [12], the output of the CUT is compressed using another LFSR, while in [13] a nonlinear finite state machine is used to obtain the ones counting in the output sequence. When both circuits operate in their test mode, a sequence of L test patterns are generated and applied to the CUT. For the ones counting methodology, the ones count in the output sequence is recorded and then compared with the pre-computed ones count of the fault-free circuit using the same L input patterns. The difference, D_i, is stored. For the signature analysis technique, the final signature is compared to the pre-computed correct signature. If they are the same the test is said to pass; otherwise it fails. This process of applying L patterns and determining the value of D_i, or pass/fail, is referred to as a *test session*.

The process for estimating error-rate consists of carrying out S test sessions, each using a different subset of test patterns from the space of all test patterns. In the following, we will concentrate on the ones counting technique. (The analogous yet radically different analysis for signature analysis is given in [12].) Carrying out these S test sessions results in S differences, denoted by D_1, D_2,…, D_s. It is assumed that all test patterns are randomly generated and that the total number of patterns, namely SxL, is a small fraction of the input pattern space. Thus these S numbers are independent and have the same distribution. The sample mean M_D and sample variance V_D are given by the equations

$$M_D = \frac{1}{S}\left(\sum_{i=1}^{S} D_i\right) \text{ and } V_D = \frac{1}{S-1}\sum_{i=1}^{S}(D_i - M_D)^2 .$$

Based on the sample mean and variance, the error-rate estimator, $\hat{r}$, given in [13] is

$$\hat{r} = 1 - \sqrt{1 - \frac{M_D^2}{L^2} - \frac{2V_D}{L}\cdot\frac{N-1}{N-L}} \qquad (1)$$

where $N=2^n$, and n is the number of input pins of the CUT. Again, $N \gg L$.

The selection of L and S depends on what objective is to be satisfied. In [13], the authors analyzed the problem of using this ones counting technique to classify chips according to their estimated error-rate.

Let r_{th} represent an error-rate threshold value, usually specified by an application, ε a margin of error allowed in the value of an estimated error-rate, and γ the confidence associated with this estimation. If the estimated error-rate, $\hat{r}$, of a CUT is smaller than r_{th}, the CUT is accepted, otherwise it is rejected. The following requirements are set for the estimation of error-rate. For a CUT with actual error-rate r,

a. If $r < r_{th}(1-\varepsilon)$, the probability of accepting the CUT is greater than or equal to γ. That is Prob($\hat{r} < r_{th}$) $\geq \gamma$ if $r < r_{th}(1-\varepsilon)$.

b. If $r > r_{th}(1+\varepsilon)$, the probability of rejecting the CUT is greater than or equal to γ. That is Prob($\hat{r} > r_{th}$) $\geq \gamma$ if $r > r_{th}(1+\varepsilon)$.

c. If $r_{th}(1-\varepsilon) \leq r \leq r_{th}(1+\varepsilon)$, the CUT may be accepted or rejected.

In [13] the authors discussed lower bounds for L and S such that the above requirements are satisfied. Their analysis showed that the lower bound for the number of test patterns per test session, L, is $\dfrac{1}{2(1-\gamma)(r_{th}\varepsilon)^2}$, and the lower bound for the number of test sessions, S, is $\dfrac{1}{2(1-\gamma)(r_{th}\varepsilon + r_{th}^2\varepsilon^2/2 + r_{th}^2\varepsilon)^2}$. Assume r_{th}=0.01, ε=0.1 and γ=0.9772. The lower bound of L is about 2.19E+7 and the lower bound of S is about 2.23E+7. Because the ones counting technique requires knowledge of the ones counting for the fault-free circuit in order to compute D_i, S values must be stored, which results in an exorbitant amount of overhead.

Employing the signature analysis technique for the same requirements as above, only 111 test patterns are needed for each of 584 test sessions. These results are counter-intuitive from the following point of view. In signature analysis one only gets a binary decision from each test session, namely pass or fail. Ignoring the issue of aliasing, this decision is always correct, but the fail can be caused by 1, 2, … up to L erroneous responses. Hence a great deal of information is lost per failing test session. On the other hand, for ones counting, one determines an actual observed ones counting, O_i, of 0, 1, 2, … L. Then the recorded value D_i is $O_i - O_i^*$, that can take on the values $-L$, $-(L-1)$, …,-1, 0, 1, 2, 3, …, L, where O_i^* is the ones counting for the fault-free circuit. Hence there appears to be a much larger amount of information in D_i than in pass/fail. In fact, if $D_i \neq 0$, then for sure the test has failed. But if $D_i = 0$, the test may have passed or failed. Thus aliasing plays a big role in ones counting. However, it seems like the value of getting a non-binary result should more than compensate for the problem of aliasing. In addition, if $L=1$ then no aliasing is possible, so if $D_i = 0$ the response is correct, and if $D_i \neq 0$ it is incorrect. Hence, for a large enough value of S, the error-rate can be accurately estimated, and thus the lower bound on L is 1. Thus it is useful to reexamine the analysis of error-rate estimation via ones counting.

To more fully characterize the statistical properties of the estimator $\hat{r}$, its mean and variance should be computed. The mean of the estimator will be close to the true error-rate when the estimator is biased, and equal to the true error-rate when the estimator is unbiased. The variance of the estimator is not directly calculated in [13]. Instead, it is implied to be equal to the variance of another random variable (namely S_e in [13]), which is not always true.

In this report, we present an analysis of the ones counting technique that is different from the analysis provided in [13]. Our analysis more accurately describes the characteristics of

ones counting technique for the purpose of error-rate estimation. We also compare this ones counting technique with the previously published work using signature analysis. The comparison shows that this ones counting technique is able to estimate error-rate as effectively as the signature analysis technique, and the hardware overheads for both techniques are comparable.

III. STATISTICAL ANALYSIS

Consider a single output combinational circuit, C, and a faulty version of this circuit, C_f. In response to each input pattern, the output of C_f can be classified into one of four types, namely, 1/1, 0/1, 0/0 and 1/0, where 1/1 means that the output of both C and C_f are 1; 0/0 means that the output of both C and C_f are 0; 0/1 means the that the output of C is 0 and C_f is 1; and 1/0 means that the output of C is 1 and C_f is 0. If all possible N input patterns are applied to C_f, a sequence of N outputs consisting of the four types of outputs is generated. The error-rate r of C_f is the ratio of the number of outputs of type 0/1 plus type 1/0 to N.

During a single test session, the observed ones count is equal to the number of type 1/1 and 0/1 outputs (type x/1). The result is subtracted by the correct ones counting of C, which is equal to the sum of type 1/1 and 1/0 outputs of C_f. This difference, denoted by D, represents the difference between the ones count generated by C_f and the ones count generated by C. It also equals the number of type 0/1 outputs minus the number of type 1/0 outputs of C_f. A complete test consists of S test sessions. Thus, recorded are S numbers, each of which is the value of D for a test session. The error-rate is estimated according to these S numbers.

Imagine that N possible outputs define a collection having four types of symbols. In each test session, we choose L outputs without replacement from the collection. From the selected outputs, we do not ascertain the number of type 0/1 or 1/0 outputs, but the difference between the number of 0/1 and 1/0 outputs. After a test session is finished, we put these outputs back into the collection. S test sessions are carried out, resulting in S numbers. From these S numbers, we estimate the fraction of outputs in the collection that are either of type 0/1 or 1/0.

Let p_1 be the fraction of 0/1 outputs in the collection, and p_2 the fraction of 1/0 outputs in the collection. Thus, $r = p_1 + p_2$. p_1, p_2 and r are all positive and less than 1. An oracle knows p_1 and p_2. We wish to estimate p_1 and p_2 since the estimated error-rate equals the sum of the estimated values of p_1 and p_2.

Assume the L outputs are drawn one at a time, and we have a counter, initialized to 0. If a type 0/1 output is drawn, the value of the counter is increased by 1; if a type 1/0 output is drawn, the value of the counter is decreased by 1; if a type 1/1 or 0/0 output is drawn, the state of the counter is not changed. The final state after L outputs are chosen, i.e., a test session, is just the state of the counter, say D. Because outputs are drawn

randomly, D is a random variable.

In the above process, it is implied that we are sampling *without replacement*. Because N is assumed to be large with respect to L, which is usually the case in practice, the change in the fraction of each type of output in the remaining collection after each of the L outputs is selected is very small and can be ignored. So we will treat this process as sampling *with replacement*. Thus, for each drawing, the probability that the counter increases by 1 is p_1, the fraction of 0/1 outputs; the probability that it decreases by 1 is p_2, the fraction of 1/0 output; and the probability that it does not change is $1 - p_1 - p_2$. Let X be a random variable such that $X = 1$ with probability p_1, $X = -1$ with probability p_2, and $X = 0$ with probability $1 - p_1 - p_2$.

Thus $D = X_1 + X_2 + \ldots + X_L$, where X_1, X_2, …, X_L are identically and independently distributed (i.i.d.) random variables with the same distribution as X. From the probability density function (PDF) of X, we see that

Expectation: $E\{X\} = p_1 - p_2$

Variance: $Var\{X\} = p_1 + p_2 - (p_1 - p_2)^2$.

Thus the expectation and variance of D are

Expectation: $E\{D\} = E\{X_1 + X_2 + \ldots + X_L\}$
$$= L \cdot E\{X\} = L(p_1 - p_2) \qquad (2)$$

Variance: $Var\{D\} = Var\{X_1 + X_2 + \ldots + X_L\}$
$$= L \cdot Var\{X\} = L(p_1 + p_2 - (p_1 - p_2)^2) \quad (3)$$

From the test process we obtain S samples of the random variable D, namely D_1, D_2, …, D_S. The sample mean M_D and the sample variance V_D are defined by the equations

$$M_D = \frac{1}{S}\sum_{i=1}^{S} D_i \text{ and } V_D = \frac{1}{S-1}\sum_{i=1}^{S}(D_i - M_D)^2 .$$

We intend to estimate the two parameters, p_1 and p_2, in the distribution of D. A generic method for building an estimator is based on approximation of the moments of the random variable [14]. The first order moment of a random variable is its expectation, which is approximated by the sample mean. The second order moment is its variance, which is approximated by the sample variance. Thus, we have

$$E\{D\} = L(p_1 - p_2) \approx M_D \qquad (4)$$

and

$$Var\{D\} = L(p_1 + p_2 - (p_1 - p_2)^2) \approx V_D . \qquad (5)$$

From (4) and (5) we solve for p_1 and p_2, and obtain

$$p_1 \approx \frac{M_D}{2L} + \frac{V_D}{2L} + \frac{M_D^2}{2L^2} \quad \text{and} \quad p_2 \approx -\frac{M_D}{2L} + \frac{V_D}{2L} + \frac{M_D^2}{2L^2} ,$$

and hence $p_1 + p_2 \approx \frac{V_D}{L} + \frac{M_D^2}{L^2}$. Because the error-rate is

equal to the sum of p_1 and p_2, the estimated error-rate, $\hat{r}$, is

$$\hat{r} = \frac{V_D}{L} + \frac{M_D^{\,2}}{L^2}, \tag{6}$$

which is also called the *estimator of error-rate*. $\hat{r}$ is a function of S samples of D. So the estimator itself is a random variable. To evaluate the performance of an estimator, its expectation and variance need to be computed. If the expectation is equal to the true value of the estimated quantity, the estimator is unbiased; otherwise it is biased. If the estimator is biased, the difference between the expectation of the estimator and the true value of the estimated quantity is of interest. Smaller differences imply better estimators. The variance of the estimator represents how close the estimation is to the expectation of the estimator. A large variance means the PDF is somewhat flat and the estimation result is likely to be poor. A small variance implies that the PDF is narrow around its expectation, and the estimation result is likely to be close to the expectation of the estimator. With the expectation and variance of the estimator, we are able to approximate the PDF of the estimator. If the type of distribution of the estimator is known, the PDF of the estimator can be expressed explicitly. If the type of distribution is unknown, as it is for the error-rate estimator, the PDF of the estimator is usually approximated by a normal distribution, which is a function of the expectation and variance of the estimator.

The procedure for deriving the expectation and the variance of the estimator $\hat{r}$ is given in Appendix A. The expectation of the estimator is

$$E\{\hat{r}\} = E\{\frac{V_D}{L} + \frac{M_D^{\,2}}{L^2}\} = (p_1 + p_2) + [(p_1 + p_2) - (p_1 - p_2)^2]/SL \tag{7}$$

and its variance is

$$Var\{\hat{r}\} = Var\{V_D/L + M_D^{\,2}/L^2\}$$
$$= \frac{2\alpha_2^2}{S} + \frac{3\alpha_2^2}{S^2} + \frac{\alpha_4 - 3\alpha_2^2 + 4\alpha_2\alpha_0^2 + 2\alpha_0\alpha_3}{LS} + \frac{2\alpha_0\alpha_3}{LS^2}$$
$$+ \frac{4\alpha_0\alpha_3 + \alpha_4 - \alpha_2^2}{L^2 S^2} - \frac{3\alpha_2^2}{LS^3} + \frac{\alpha_4 - 3\alpha_2^2}{L^2 S^3} + \frac{\alpha_4 - 3\alpha_2^2}{L^3 S^3} \tag{8}$$

where $\quad \alpha_0 = p_1 - p_2, \qquad\qquad \alpha_2 = p_1 + p_2 - \alpha_0^2,$

$$\alpha_3 = \alpha_0 - 3(p_1 + p_2)\alpha_0 + 2\alpha_0^3 \qquad\qquad \text{and}$$

$$\alpha_4 = (p_1 + p_2) + (6(p_1 + p_2) - 4)\alpha_0^2 - 3\alpha_0^4.$$

The true value of the error-rate r to be estimated is $p_1 + p_2$. Equation (7) shows that the estimator is biased. However, when SL is large, the terms with SL can be ignored and the estimator becomes unbiased. Later we show that for the problems addressed in this report, SL is large, and for these cases $E\{\hat{r}\} \approx p_1 + p_2 = r$.

We use the normal distribution $N_{\hat{r}}(E\{\hat{r}\}, Var\{\hat{r}\})$ to approximate the distribution of the estimator. Thus the PDF of the estimator can be expressed as

$$P_{\hat{r}}(\hat{r}) = \frac{1}{\sqrt{2\pi Var\{\hat{r}\}}} e^{-\frac{(\hat{r}-r)^2}{2Var\{\hat{r}\}}}. \tag{9}$$

We are interested in having the estimated error-rate be within a certain range of accuracy, say $[\,r(1-\varepsilon)\,,\,r(1+\varepsilon)\,]$, with confidence not less than γ, where $0 < \varepsilon \ll 1$ and γ is between 0 and 1. Thus, we require

$$\gamma \leq \int_{r(1-\varepsilon)}^{r(1+\varepsilon)} P_{\hat{r}}(\hat{r}) d\hat{r} = \int_{r(1-\varepsilon)}^{r(1+\varepsilon)} \frac{1}{\sqrt{2\pi Var\{\hat{r}\}}} e^{-\frac{(\hat{r}-r)^2}{2Var(\hat{r})}} d\hat{r}$$

$$= \int_{-r\varepsilon/\sqrt{Var\{\hat{r}\}}}^{r\varepsilon/\sqrt{Var\{\hat{r}\}}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

$$= 1 - 2Q\left(r\varepsilon/\sqrt{Var(\hat{r})}\right) \tag{10}$$

where the function Q is defined as $Q(x) = \int_x^\infty e^{-t^2/2}/\sqrt{2\pi}\,dt$

(see Appendix B)

Changing the form of (10), we have

$$Q^{-1}\left((1-\gamma)/2\right) \leq r\varepsilon/\sqrt{Var\{\hat{r}\}} \tag{11}$$

$Var\{\hat{r}\}$ is a function of r ($=p1+p2$), $p_1 - p_2$, S and L. Thus, (11) includes six quantities, namely ε, γ, r, $p_1 - p_2$, S and L, where S and L are unknown and ε and γ are given as part of the test specifications. To determine values for S and L, which are test parameters for carrying out error-rate estimation, some additional constraints and/or objective functions are needed. Referring back to our VLSI test problem, some quantities of interests are listed next.

a) Minimize $S{\times}L$, which primarily determines the total test time;

b) Minimize S, which primarily determines the storage cost for the correct ones counting;

c) Minimize $c_1 S{\times}L + c_2 S{\times}L$, which is the weighted cost for both test time and storage cost, where c_1 and c_2 are both non-negative cost coefficient.

Referring back to (11), r is the true error-rate only known to an oracle, but we can guess a value for r and refine our guess once we have a value for the estimator. The quantity $(p_1 - p_2)$ is also unknown, but again we can attempt to approximate it. Because the approximations are different for various situations, we will deal with these issues in the next section entitled case studies.

IV. CASE STUDIES

In [13], the number of test patterns per session, L, has a lower bound. However, our analysis shows that any positive integer is feasible for L. First we consider two cases, namely $L=1$ and $1/L\approx 0$ (i.e., L is very large). From (11) and (8) we see that S and L are inversely proportional to each other. So $L=1$ results in an upper bound for S, and $1/L\approx 0$ results in a lower bound of S. Then we consider the symmetric case of $p_1 = p_2 > 0$. At last we consider the general case. In the following (except in Section 4.4), we assume $\gamma = 0.9972$ and hence

$Q^{-1}((1-\gamma)/2)=3$. Then (11) reduces to

$$Var\{\hat{r}\} \leq r^2 \varepsilon^2 / 9 \qquad (12)$$

4.1 Case 1: L=1

From (8), we have $Var\{\hat{r}\} = (2\alpha_4 - 9\alpha_2^2)/S^3 + (2\alpha_2^2 + 6\alpha_0\alpha_3 + \alpha_4)/S^2 + (\alpha_4 - \alpha_2^2 + 4\alpha_2\alpha_0^2 + 2\alpha_0\alpha_3)/S$ with $L=1$. As $\alpha_0 = p_1 - p_2$, we have $0 \leq \alpha_0^4 = (p_1 - p_2)^4 \leq \alpha_0^2 = (p_1 - p_2)^2 \leq |\alpha_0| = |p_1 - p_2| \leq p_1 + p_2 = r \leq 1$. From the definitions of α_2, α_3 and α_4, we know that all of these terms are of the order of r. So $\alpha_4 - \alpha_2^2 + 4\alpha_2\alpha_0^2 + 2\alpha_0\alpha_3$, $2\alpha_2^2 + 6\alpha_0\alpha_3 + \alpha_4$ and $2\alpha_4 - 9\alpha_2^2$ are all of the order of r. Keep terms with S in the denominator and ignore terms with higher orders of S, we have

$$Var\{\hat{r}\} = \frac{\alpha_4 - \alpha_2^2 + 4\alpha_2\alpha_0^2 + 2\alpha_0\alpha_3}{S} \qquad (13)$$

and (12) can be rewritten as

$$\frac{\alpha_4 - \alpha_2^2 + 4\alpha_2\alpha_0^2 + 2\alpha_0\alpha_3}{S} \leq r^2 \varepsilon^2 / 9. \qquad (14)$$

Replacing α_2, α_3 and α_4 with their functions of r in (14) leads to

$$S \geq \frac{9}{r^2 \varepsilon^2}\left(r - r^2 + (6r - 2)\alpha_0^2 - 4\alpha_0^4\right). \qquad (15)$$

Since $\alpha_0 (= p_1 - p_2)$ and r are unknown, we cannot choose the value S to be $9\left(r - r^2 + (6r - 2)\alpha_0^2 - 4\alpha_0^4\right)/(r^2\varepsilon^2)$. However, we can choose S to be the maximum of $9\left(r - r^2 + (6r - 2)\alpha_0^2 - 4\alpha_0^4\right)/(r^2\varepsilon^2)$. Thus, (15) is satisfied and it is guarantees that the estimated error-rate is in the range of $[r(1-\varepsilon), r(1+\varepsilon)]$ with confidence γ.

For $r \leq 1/3$, which is typically the case, $\alpha_0^2 = 0$ maximizes $r - r^2 + (6r - 2)\alpha_0^2 - 4\alpha_0^4$. So we choose S to be

$$\left. \frac{9}{r^2\varepsilon^2}\left(r - r^2 + (6r - 2)\alpha_0^2 - 4\alpha_0^4\right)\right|_{\alpha_0^2 = 0}, \text{ i.e.,}$$

$$S^{C1} = \frac{9(1-r)}{r\varepsilon^2}. \qquad (16)$$

For example, for $r=0.01$, $\varepsilon=0.05$ and $L=1$, we choose S to be 3.6E+5. When $1/3<r<1$, $\alpha_0^2 = (3r-1)/4$ maximizes $r - r^2 + (6r - 2)\alpha_0^2 - 4\alpha_0^4$. So S can be chosen as

$$\left. \frac{9}{r^2\varepsilon^2}\left(r - r^2 + (6r - 2)\alpha_0^2 - 4\alpha_0^4\right)\right|_{\alpha_0^2 = (3r-1)/4}, \text{ i.e.,}$$

$\frac{9(5r^2 - 2r + 1)}{4r^2\varepsilon^2}$. For $r=0.4$, $\varepsilon=0.05$ and $L=1$, we choose S to be 5625.

4.2 Case 2: 1/L ≈ 0

For the case of $1/L \approx 0$, not only do we assume that L is very large, but also that $L >> S$, in which case we ignore all terms in (8) with an L, and obtain

$$Var\{\hat{r}\} = \frac{2\alpha_2^2}{S} + \frac{3\alpha_2^2}{S^2}. \qquad (17)$$

Ignoring the term with S^2 and replacing α_2 with its function of r and α_0, (17) reduces to $Var\{\hat{r}\} = 2(r - \alpha_0^2)^2 / S$. From (12) and (17), we have

$$S \geq \frac{18(r - \alpha_0^2)^2}{r^2\varepsilon^2}. \qquad (18)$$

Similar to case 1, we can choose the value of S to be the maximum of $18(r - \alpha_0^2)^2/r^2\varepsilon^2$. Since the maximum value of $18(r - \alpha_0^2)^2/r^2\varepsilon^2$ is $18/\varepsilon^2$, we choose S as

$$S^{C2} = 18/\varepsilon^2. \qquad (19)$$

For example, for $\varepsilon=0.05$, we get $S^{C2} = 7200$ and thus $L >> 7200$. From (19) we see that when L is very large, the number of test sessions required in error-rate estimation is independent of the error-rate, and only depends on the accuracy and confidence of the estimation. This follows since, if the error-rate is extremely small, then the allowable error in our estimation as specified by ε allows our test methodology to work, and for larger values of error-rate, there is enough information gathered by using these values of L and S^{C2} to again estimate the error-rate.

If we choose $L=7200$ and solve for S using (19), we get $L=S$. Then the condition to approximate (8) with (17) is not satisfied. In this case, we cannot use (19) to select S. Case 1 and Case 2 represent two extreme values of L resulting in upper and lower bounds for S.

4.3 Case 3: $p_1 = p_2 > 0$

When $p_1 = p_2$, the probability of observing a type 0/1 or 1/0 output are the same. Thus, the expected value of D for each session is zero, i.e., the sample mean of D, M_D, is close to zero. In this case, it appears that the ones counting test methodology becomes ineffective. However, when $M_D = 0$ the estimator (6) reduces to V_D/L. This means that the error-rate can be derived solely from the variance of D. With $p_1 = p_2$, we have $\alpha_0 = \alpha_3 = 0$ and $\alpha_2 = \alpha_4 = r$. From (8), we have

$$Var\{\hat{r}\} = \frac{2r^2}{S} + \frac{3r^2}{S^2} + \frac{r - 3r^2}{LS} + \frac{r - r^2}{L^2S^2}$$

$$-\frac{3r^2}{LS^3}+\frac{r-3r^2}{L^2S^3}+\frac{r-3r^2}{L^3S^3}. \tag{20}$$

Assume S is large. When compared with $2r^2/S$, $3r^2/S^2$ can be ignored. When compared with $(r-3r^2)/LS$, we can ignore $3r^2/LS^3$, $(r-r^2)/L^2S^2$, $(r-3r^2)/L^2S^3$ and $(r-3r^2)/L^3S^3$. Thus, (20) reduces to

$$Var\{\hat{r}\}=\left(2r^2+(r-3r^2)/L\right)\big/S. \tag{21}$$

From (12) and (21), we have $\left(2r^2+(r-3r^2)/L\right)\big/S \le r^2\varepsilon^2/9$, i.e.,

$$S^{C3}\ge\frac{9}{\varepsilon^2}\left(2+\frac{1}{L}(\frac{1}{r}-3)\right) \tag{22}$$

Any positive integer is legitimate for L. When $L=1$, $S=9(1-r)/(r\varepsilon^2)$. When L is very large, $S=18/\varepsilon^2$. These results for S are consistent with those derived for Cases 1 and 2. Fig. 2 shows the relationship between S and L for different error-rates and $\varepsilon=0.05$.
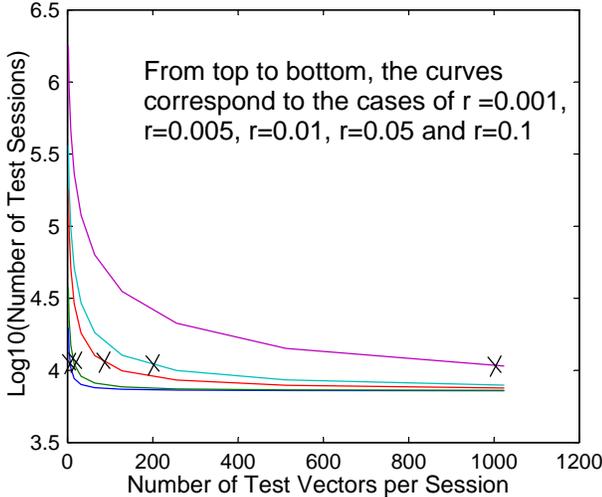
Fig. 2. S vs. L based on Eq. (22) for different error-rates and $\varepsilon=0.05$. The points marked by an 'X' correspond to the selection of S and L that minimize SxL.

Consider the case of a single stuck at fault in a single-output XOR circuit that causes half of all outputs to be wrong. Among those erroneous outputs, half are of type 0/1 and the other half of type 1/0. So $p_1=p_2=1/4$ and $r=1/2$. This provides an example of Case 3. If the output line is stuck-at 0, then again $r=1/2$ but now all errors are of type 1/0.

To see how the variance is instrumental in determining the estimated value of error-rate, again consider a circuit such that a randomly selected half of all possible input patterns map into 1, and the other half map into 0. Now consider a faulty version of this circuit, where again a randomly selected half of all possible input patterns map into 1, and the other half map into 0. So the true error-rate of the faulty circuit is 1/2. Now, for both the good and faulty circuits, the average value of the ones counting is $L/2$. Thus, the expected value of D (the difference of ones counting between good circuit and faulty circuit) is zero. In addition, for the faulty circuit, the

probability of observing a 0/0, 0/1, 1/0 or 1/1 type response is 1/4, i.e., $p_1=p_2=1/4$. Assume $L=4$. For each session, the possible values of D are –4, –3, –2, –1, 0, 1, 2, 3 and 4. If D=–4 for a session, then all four outputs are 1/0 type. Thus, the probability of D=–4 is $(1/4)^4$= 1/256. If D=–3 for a session, then three outputs are 1/0 type and one output is either 0/0 or 1/1 type. Thus, the probability of D=–3 is $4\times(1/4)^3\times(1/2)$= 1/32. Similarly, we can computer the probability of D being –2, –1, 0, 1, 2, 3 or 4. As a result, the probabilities of D being –4, –3, –2, –1, 0, 1, 2, 3 and 4 are 1/256, 1/32, 7/64, 7/32, 35/128, 7/32, 7/64, 1/32 and 1/256, respectively. With the distribution function of D, we have the expectation of D to be zero and the variance to be 2. From the expectation and variance of D, we know that the sample mean of D, M_D, is about zero and the sample variance of D, V_D, is about 2. From the estimator $V_D/L+M_D^2/L^2$, we obtain the estimated error-rate to be 1/2, which matches the true error-rate.

4.4 General Case

For the general case, we make no assumptions of the values of p_1, p_2 and γ. However, we assume S is large. To make our discussion clear, we show a copy of (8) below.

$$Var\{\hat{r}\}=Var\{V_D/L+M_D^2/L^2\}$$
$$=\frac{2\alpha_2^2}{S}+\frac{3\alpha_2^2}{S^2}+\frac{\alpha_4-3\alpha_2^2+4\alpha_2\alpha_0^2+2\alpha_0\alpha_3}{LS}+\frac{2\alpha_0\alpha_3}{LS^2}$$
$$+\frac{4\alpha_0\alpha_3+\alpha_4-\alpha_2^2}{L^2S^2}-\frac{3\alpha_2^2}{LS^3}+\frac{\alpha_4-3\alpha_2^2}{L^2S^3}+\frac{\alpha_4-3\alpha_2^2}{L^3S^3}$$

Because S is large, the term $3a_2^2/S^2$ can be ignored when compared to $2a_2^2/S$. For the same reason, the terms $2\alpha_0\alpha_3/LS^2$, $(4\alpha_0\alpha_3+\alpha_4-\alpha_2^2)/L^2S^2$, $3a_2^2/LS^3$, $(\alpha_4-3a_2^2)/L^2S^3$ and $(\alpha_4-3a_2^2)/L^3S^3$ can be ignored when compared to $(\alpha_4-3\alpha_2^2+4\alpha_2\alpha_0^2+2\alpha_0\alpha_3)/LS$. Thus, the variance of the estimator becomes

$$Var\{\hat{r}\}=\frac{2\alpha_2^2}{S}+\frac{\alpha_4-3\alpha_2^2+4\alpha_2\alpha_0^2+2\alpha_0\alpha_3}{LS}. \tag{23}$$

Replacing α_2, α_3 and α_4 with their functions of r and α_0, we have

$$Var\{\hat{r}\}=\frac{2(r-\alpha_0^2)^2}{S}+\frac{(r-3r^2)+(10r-2)\alpha_0^2-6\alpha_0^4}{LS}. \tag{24}$$

Then, (11) becomes

$$\frac{2(r-\alpha_0^2)^2}{S}+\frac{(r-3r^2)+(10r-2)\alpha_0^2-6\alpha_0^4}{LS}\le\frac{r^2\varepsilon^2}{\left(Q^{-1}((1-\gamma)/2)\right)^2}$$

and hence

$$S \geq \frac{\left(Q^{-1}((1-\gamma)/2)\right)^2}{r^2\varepsilon^2}\left(2(r-\alpha_0^2)^2 + \frac{(r-3r^2)+(10r-2)\alpha_0^2-6\alpha_0^4}{L}\right). \quad (25)$$

Without knowing the value of α_0, we choose S to be the maximum achievable value of right hand side of (25). It can be shown (see Appendix C) that when r is less than 1/5, which is generally the case of interest, $\alpha_0 = 0$ results in (25) being maximal. Thus we choose S as

$$S = \frac{\left(Q^{-1}((1-\gamma)/2)\right)^2}{\varepsilon^2}\left(2+\frac{1}{L}\left(\frac{1}{r}-3\right)\right). \quad (26)$$

For $\gamma = 0.9972$, (26) reduces to (22). This is expected because Case 3, from which (22) is derived, is a special case of Case 4. For $L=1$, (26) reduces to (16). For $1/L\approx0$, (26) reduces to (19). These results imply that the analysis of the general case is consistent with the analysis of its special cases.

Setting $L= \infty$ allows us to find a lower bound on S. For $\gamma=0.9972$, this lower bound is given by (19). For $\varepsilon=0.1$, the lower bound is 1800. Usually ε is much smaller than 0.1 and, as ε decreases the lower bound on S increases. Thus in (8), where an S and LS term exist, it is appropriate to ignore terms containing S^2, LS^2, L^2S^2, LS^3, L^2S^3 and L^3S^3. This justifies the approximations used to obtain (24) from (8).

Now consider minimizing test time which is proportional to SxL. Assume $\gamma=0.9972$. From (26), we have $SL = \frac{9}{\varepsilon^2}\left(2L+\frac{1}{r}-3\right)$. When $2L = 1/r$, then $S = 9(4-6r)/\varepsilon^2$ and SxL has a minimal value of $9(2/r-3)/\varepsilon^2$. When r is small, $S \approx 36/\varepsilon^2$ and $(SxL)_{min} \approx 18/(r\varepsilon^2)$. In Fig. 2, the points marked by 'X' correspond to the selections of S and L that minimize SxL. It can be seen that for different values of error-rate, the values of S are almost the same, namely $36/\varepsilon^2$.

For error-rate estimation using signature analysis [12], it is recommended to set $L = 1/r$, which leads to $S = 15/\varepsilon^2$ and $SL = 15/(r\varepsilon^2)$. Thus we see that the ones counting technique for error-rate estimation is comparable to signature analysis in terms of total test time, and a little higher in terms of overhead cost.

V. Simulation

In the above analysis, we approximated the probability density function of the estimator with the normal distribution $N_{\hat{r}}(E\{\hat{r}\}, Var\{\hat{r}\})$. Then we developed a way to select S and L to satisfy the accuracy and confidence requirement of error-rate estimation based on this approximation. In this section, we describe our results of estimating the error-rate via simulation. By repeating the simulation process many times, we can collect a large number of estimated error-rates and compare their distribution with $N_{\hat{r}}(E\{\hat{r}\}, Var\{\hat{r}\})$.

The simulation is implemented as follows.

1) A random number generator generates three numbers, namely a 1 with probability p_1, −1 with probability p_2 and 0 with probability $1- p_1-p_2$.
2) The number of 1s and the number of −1s in a sequence of L generated numbers are counted separately. Their difference is recorded. This represents a test session.
3) S test sessions are carried out, and S numbers are recorded. With the estimator presented in Section 3, the error-rate $p_1+ p_2$ is estimated.
4) The above procedure is repeated 3000 times and results in 3000 error-rate estimations. The distribution of the estimated error-rates is generated.

We use the MATLAB tool "normplot" to determine if this data is consistent with a normal distribution. Normplot displays the cumulative distribution of the data. In the plot, a superimposed line is drawn to fit the sample data. If the data are normally distributed, the plot appears linear.

First, we set $p_1=0.006$, $p_2=0.004$, $L=1$ and $S=1.0E+5$. Fig. 3a shows the distribution of the data and it appears to be
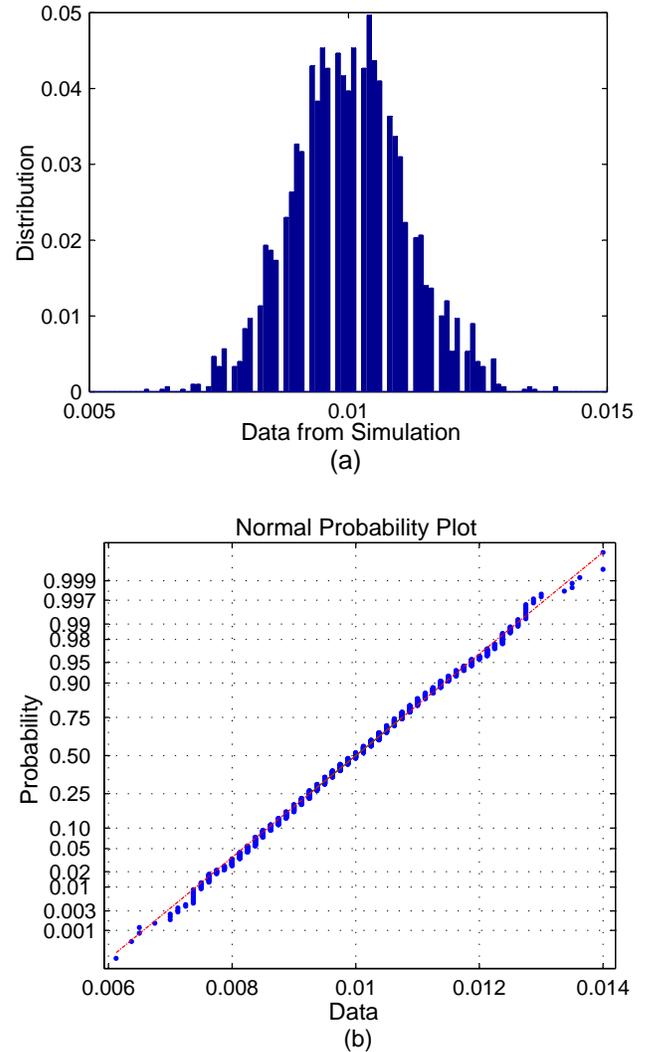
Fig. 3. (a) Distribution of estimated error-rate data from simulation. (b) The output from MATLAB tool "normplot" to test whether the data are normally distributed. For this figure, $p_1=0.006$, $p_2=0.004$, $L=1$ and $S=1.0E+5$.

normal. The output of "normplot" is shown in Fig. 3b, and is fairly linear, confirming that the data has a normal distribution. Now, a normal distribution can be defined by its mean and standard deviation. From the data, we estimate the mean to be 0.01 and the variance to be 1.21E-6. So the set of error-rate data has normal distribution N(0.01, 1.21E-4). In Section 3, we used a normal distribution to approximate the distribution of the estimator. The mean of the estimator is given by (7), and results in a value of 0.01. The variance of the estimator is given by (8), and results in a value of 1.23E-6. So in our analysis, we would use N(0.01, 1.23E-6) to be the distribution of the estimator. Thus our analytical results closely match the simulation results.

Next, we choose different values of L and S while keeping p_1=0.006 and p_2=0.004. Fig. 4 shows the distribution of the estimated error-rate data from simulation and the normal distribution test from "normplot" for the case where L=1000 and S=2000. L =1000 is about 10 times $1/r$. For ε=0.1, the lower bound on S is 1800. S =2000 is close to the lower bound on S, and thus the case outlined in Section 4.2 is satisfied. The

figure shows that the simulation data are normally distributed. From the data, we estimate its normal distribution to be N(0.01, 1.05E-7). From the analysis in Section 3, we obtain the distribution of the estimator to be N(0.01, 1.048E-7), which again is an excellent match.

Finally, we consider the case where p_1= p_2=0.005, L=50 and S=4000. Since L is small, one would expect that for most test sessions few if any errors would occur. In this case, the estimated error-rate is mainly derived on the variance of the sample data as mentioned in Section 4.3. The distribution of error-rate data from simulation and the output of "normplot" are displayed in Fig. 5, which shows the simulation data has a normal distribution. The distribution function of the data is estimated to be N(0.01, 9.85E-8), which matches the distribution of the estimator from analysis, namely N(0.01, 9.85E-8).

VI. CLASSIFYING CHIPS VIA THEIR ERROR-RATE

One application for error-rate estimation is to assign chips

Fig. 4. (a) Distribution of estimated error-rate data from simulation. (b) Output from MATLAB tool "normplot" to test whether data are normally distributed. For this figure, p_1=0.006, p_2=0.004, L=1000 and S=2000.
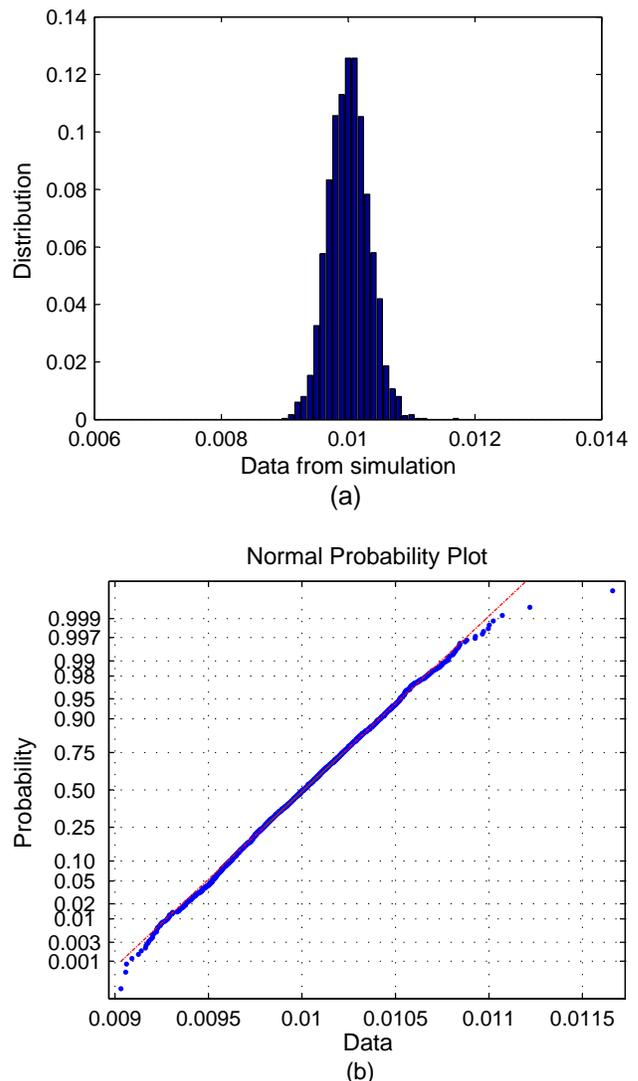
Fig. 5. (a) Distribution of estimated error-rate data from simulation. (b) Output from MATLAB tool "normplot" to test whether data are normally distributed. For this figure, p_1=0.005, p_2=0.005, L=50 and S=4000.

to bins that correspond to error-rate ranges that are defined by threshold error-rate values. Recall that a threshold separates a range into two adjacent sub-ranges. Threshold r_{th} divides the domain of error-rates into two range $(0, r_{th})$ and $(r_{th}, 1)$, and consequently failing chips are partitioned into two types, A and B. The error-rate of a type A chip is equal to or less than r_{th}, while that of a type B is greater than r_{th}. Testing classifies a chip into type A or type B according to the estimated error-rate $\hat{r}$. Namely, if $\hat{r} < r_{th}$, the chip is classified as type A; otherwise, it is classified as type B. Unfortunately, the random variable $\hat{r}$ can be greater than r_{th} even though the true error-rate, r, is smaller than r_{th} and vice versa. The chance for this to occur increases rapidly as $\hat{r}$ approaches the value r_{th}. So the test can classify chips incorrectly. In statistics such a test is called a *hypothesis test*. In our case, the two hypotheses are:

H0: The chip is type B, i.e., $r > r_{th}$;

H1: The chip is type A, i.e., $r \le r_{th}$.

This test generates four possible outcomes. 1) The chip is type B, and classified as type B; 2) the chip is type B, and classified as type A; 3) the chip is type A, and classified as type A; and finally 4) the chip is type A and classified as type B. Outcome 2) results in a lower price chip sold erroneously at a higher price, while outcome 4) results in a higher price chip sold at a lower price. Outcome 4) is likely acceptable to customers and outcome 2) is not. So the test should limit the probability of the occurrence of outcome 2).

Assume it is required that the probability of any type B chip being classified as type A be smaller than β, where $\beta \ll 1$. According to the analysis of error-rate estimation, the estimated error-rate has normal distribution with its expectation being the true error-rate. Fig. 6 shows the distribution of estimated error-rate of a chip whose true error-rate is r_{th} and a chip whose true error-rate is greater than r_{th}. It can be seen that for a chip with true error-rate r greater than r_{th}, the further the true error-rate r is from r_{th}, the lower that probability of outcome 2) occurring. This probability is represented by the dash area under the curve. However, when the true error-rate is equal to r_{th}, the probability of outcome 2) is always 50%. Thus, the requirement is never satisfied. To solve this issue, we define another threshold, namely r_{thn}, that is smaller than r_{th}, and postulate the following classification criterion:
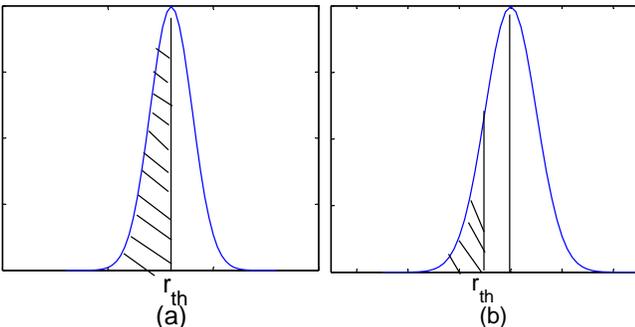
Fig. 6. (a) PDF of estimated error-rate of a chip whose true error-rate is r_{th}. (b) PDF of estimated error-rate of a chip whose true error-rate is greater than r_{th}.

If the estimated error-rate of a chip is smaller than r_{thn}, it is classified as type A. If the estimated error-rate of a chip is equal to or greater than r_{thn}, it is classified as type B.

Assume that the probability of outcome 2) occurring is still limited to β. Thus this constraint also holds when $r = r_{th}$, hence the probability of outcome 2) is smaller than β when $r > r_{th}$. So we require $\beta \ge \text{Prob}(\hat{r} < r_{thn} \mid r = r_{th})$. When $r = r_{th}$, the estimated error-rate has normal distribution $N(r_{th}, Var\{\hat{r}\})$. Thus, $\text{Prob}(\hat{r} < r_{thn} \mid r = r_{th}) =$

$$Q\left(\frac{r_{th} - r_{thn}}{\sqrt{Var\{\hat{r}\}}}\right) \le \beta, \qquad \text{which} \qquad \text{is} \qquad \text{equivalent} \qquad \text{to}$$

$$\frac{(r_{th} - r_{thn})^2}{(Q^{-1}(\beta))^2} \ge Var\{\hat{r}\}.$$

From (24), we already have

$$Var\{\hat{r}\} = \frac{2(r_{th} - \alpha_0^2)^2}{S} + \frac{(r_{th} - 3r_{th}^2) + (10r_{th} - 2)\alpha_0^2 - 6\alpha_0^4}{LS}.$$

Thus

$$\frac{(r_{th} - r_{thn})^2}{(Q^{-1}(\beta))^2} \ge \frac{2(r_{th} - \alpha_0^2)^2}{S} + \frac{(r_{th} - 3r_{th}^2) + (10r_{th} - 2)\alpha_0^2 - 6\alpha_0^4}{LS} \qquad (27)$$

or equivalently,

$$S \ge \frac{(Q^{-1}(\beta))^2}{(r_{th} - r_{thn})^2}\left(2(r_{th} - \alpha_0^2)^2 + \frac{(r_{th} - 3r_{th}^2) + (10r_{th} - 2)\alpha_0^2 - 6\alpha_0^4}{L}\right) \qquad (28)$$

(28) describes the requirement for S and L such that the probability of outcome 2) occurring is limited to β. Without knowing the value of α_0, we choose S to be greater than the maximum value of the right hand side of (28). With r_{th} being smaller than 0.2, the right hand side of (28) is maximal at the value $\frac{(Q^{-1}(\beta))^2}{(r_{th} - r_{thn})^2}\left(2r_{th}^2 + \frac{(r_{th} - 3r_{th}^2)}{L}\right)$ when $\alpha_0 = 0$. So we choose S according to the expression

$$S = \frac{(Q^{-1}(\beta))^2}{(r_{th} - r_{thn})^2}\left(2r_{th}^2 + \frac{(r_{th} - 3r_{th}^2)}{L}\right) \qquad (29)$$

In (29), β and r_{th} are given in the test specification. We need to determine the value of S, L and r_{thn}. From our previous analysis, the selection of S and L depends on what cost function is minimized. Assume L has been determined. Then S is determined by r_{thn}. As r_{thn} decreases, S decreases. Because smaller values of S result in less storage cost, it is important to keep S small. However, smaller r_{thn} causes more type A chips to be classified as type B, meaning a loss in profit. So there is a tradeoff between storage cost and profit.

For a type A chip, its true error-rate, r, may be smaller or greater than r_{thn}. Given L and S, the probability of outcome 4) can be computed. Consider the case of $r < r_{thn}$. The estimated error-rate has normal distribution $N\left(r, \frac{2(r - \alpha_0^2)^2}{S} + \frac{(r - 3r^2) + (10r - 2)\alpha_0^2 - 6\alpha_0^4}{LS}\right)$. The probability of

outcome 4), i.e. the estimated error-rate being greater than r_{thn} is

$$Q\left(\frac{r_{thn}-r}{\sqrt{2(r-\alpha_0^2)^2/S+((r-3r^2)+(10r-2)\alpha_0^2-6\alpha_0^4)/(LS)}}\right). \quad (30)$$

When $r < 0.2$ and $\alpha_0 = 0$, (30) takes on a maximal value

of $Q\left(\frac{r_{thn}-r}{\sqrt{2r^2/S+(r-3r^2)/(LS)}}\right)$, which is the upper bound of

the probability of outcome 4) for the case of $r < r_{thn}$. When $r_{thn} < r < r_{th}$, the probability of outcome 4) is

$$1-Q\left(\frac{r-r_{thn}}{\sqrt{2(r-\alpha_0^2)^2/S+((r-3r^2)+(10r-2)\alpha_0^2-6\alpha_0^4)/(LS)}}\right). \quad (31)$$

Since $|\alpha_0| = |p_1 - p_2| \le r$, $\alpha_0^2 \le r^2$, it can be shown that

if $r < 1/3$ then (31) is maximal when $\alpha_0^2 = r^2$ and has the

value $1-Q\left(\frac{r-r_{thn}}{\sqrt{2(r-r^2)^2/S+(r-5r^2+10r^3-6r^4)/(LS)}}\right)$.

Similarly, the probability of outcome 2) for the case of $r > r_{th}$ can be computed. Table 1 summarizes the upper bounds on the probability of erroneous classification for these three cases.

Similar to the error-rate estimation technique based on signature analysis [12], it is not necessary to apply all test sessions before making a decision because the farther the true error-rate away from the threshold, the less probability of making wrong classification. The number of test sessions is based on the assumption of the worst case, i.e., when $r = r_{th}$. However, a defective chip usually does not represent the worst case, and sometimes never represents such a case. The classification requirement is that the probability of an estimated error-rate less than r_{thn} be smaller than β if the true error-rate is greater than r_{th}. So it is possible to make a decision without executing all S test sessions as long as the probability of making a wrong decision is smaller than β. Let $S_{ms}(r)$ be the minimal number of test sessions required for a chip with error-rate r that satisfies the constraint imposed by β.

For $r > r_{th}$, the upper bound on the probability that the estimated error-rate is smaller than r_{thn} is

$Q\left(\frac{r-r_{thn}}{\sqrt{2r^2/S+(r-3r^2)/(LS)}}\right)$. For $r < r_{thn}$, the upper bound of

the probability that estimated error-rate is greater than r_{thn} is

$Q\left(\frac{r_{thn}-r}{\sqrt{2r^2/S+(r-3r^2)/(LS)}}\right)$. Both formulas are listed in

Table 1. Then $S_{ms}(r)$ should satisfy

$$Q\left(\frac{r-r_{thn}}{\sqrt{2r^2/S_{ms}+(r-3r^2)/(LS_{ms})}}\right)=\beta \quad \text{if } r > r_{th}, \quad (32)$$

and

$$Q\left(\frac{r_{thn}-r}{\sqrt{2r^2/S_{ms}+(r-3r^2)/(LS_{ms})}}\right)=\beta \quad \text{if } r < r_{thn}. \quad (33)$$

TABLE I
THE MAXIMAL LIKELIHOOD OF MAKING AN ERRONEOUS CLASSIFICATION

	Upper Bound of the Probability of Making Erroneous Classification
True error-rate $r<r_{thn}$, and estimated error-rate $\hat{r} > r_{thn}$ (Type A classified as type B)	$Q\left(\dfrac{r_{thn}-r}{\sqrt{2r^2/S+(r-3r^2)/(LS)}}\right)$
$r_{thn}<r<r_{th}$ and $\hat{r} > r_{thn}$. (Type A classified as type B)	$1-Q\left(\dfrac{r-r_{thn}}{\sqrt{\dfrac{2(r-r^2)^2}{S}+\dfrac{r-5r^2+10r^3-6r^4}{LS}}}\right)$
$r>r_{th}$ and $\hat{r} < r_{thn}$. (Type B classified as type A)	$Q\left(\dfrac{r-r_{thn}}{\sqrt{2r^2/S+(r-3r^2)/(LS)}}\right)$

When $r_{thn} < r < r_{th}$, $S_{ms}(r)$ is equal to S as specified by (29). As an example, let $r_{thn} = 0.009$, $r_{th} = 0.01$ and $\beta = 0.05$. L is chosen to be 100 $(=1/r_{th})$. Fig. 7 shows the minimal number of test sessions $S_{ms}(r)$ for different true error-rates r. As r moves away from the range of $[r_{thn}, r_{th}]$, $S_{ms}(r)$ quickly decreases. So for error-rates far from $[r_{thn}, r_{th}]$, only a small number of test sessions are needed.

To be able to make early decision before applying all test sessions, the test procedure must be modified. An original test is divided into multiple phases, each of which consists of a disjoint subset of the S test sessions. After each test phase is completed, the error-rate is estimated based on the results from the test sessions applied so far, and assuming the estimated error-rate is the true error-rate. If the estimated error-rate is smaller than r_{thn}, the probability of making an erroneous classification is calculated using (33). If the estimated error-rate is greater than r_{th}, the probability of making an erroneous classification is calculated using (32). If the computed probability is smaller than β, the test stops, otherwise it continues. If the estimated error-rate is in the range $[r_{thn}, r_{th}]$, the test also continues unless all test sessions
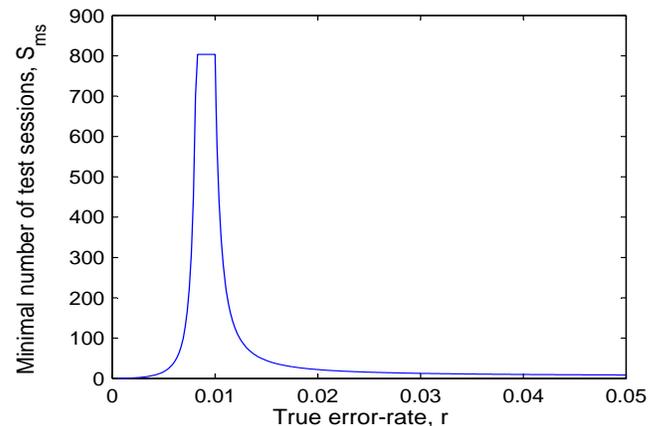
Fig.7.   The minimal number of test sessions for different true error-rates.

have been run, in which case testing is finished.

VII. Experimental Results

To validate this chip classification technique, we applied it to the ISCAS'85 benchmark circuit C432 that has 7 primary outputs. Because our technique is currently only applicable to single output circuits, we created seven single output circuits from C432, labeled as C432_1, C432_2, C432_3, C432_4, C432_5, C432_6 and C432_7. The seven circuits have the same netlist as C432. For each of these netlists, only one primary output of C432 is treated as the output of the new circuit and other outputs are treated as internal wires. For example, the output pin 1 of C432 is the output of C432_1, and output pin 2, 3, 4, 5, 6 and 7 of C432 are treated as internal wires in C432_1. To model a defect, we used the single stuck-at fault model. Each of these circuits has 864 single stuck-at faults. Thus, corresponding to each fault-free circuit, there are 864 faulty circuits. Consider the 864 faulty circuits of C432_7. Since we know the actual faults in the circuit, we can obtain their actual error-rates (see Fig. 8).

To classify these circuits, we set r_{th}=0.02, r_{thn}=0.019, β=0.05 and L=50. The error-rate of a type A circuit is in the range (0, 0.02), and for a type B circuit [0.02, 1). From (29), the maximal number of test sessions, S, is 3182. We use the multi-phase test scheme described in the previous section for classification. In the first phase, S^*=20 test sessions are executed. The collected data is not statistically meaningful if S^* is too small. In the following phases, only one test session is applied. Even though more test sessions can be applied, we choose one because we want to find the exact number of the test sessions needed. After each phase, the estimated error-rate is computed and the condition for stopping is checked. The total number of test sessions for each circuit is recorded. Fig. 9 shows the histogram of test sessions for all the circuits. From Fig. 9, it is seen that for many circuits, only a small number of test sessions are needed. This means that the error-rate for each of these circuits is far away from the range (0.019, 0.02). This result is consistent with Fig. 8, that shows that only a small fraction of the circuits have an error-rate among and near the range (0.019, 0.02).

To further demonstrate the correlation between the number of test sessions and error-rate, in Table 2 we list the average number of test sessions for different sub-ranges of error-rate. For error-rate far from 0.019~0.02, the average number of test sessions is small. As the error-rate gets close to 0.019~0.02, the average number of test sessions increases. This is consistent with the analysis in Section 6. For error-rates in the range 0.019~0.02, the average number of test sessions is 995.44, which is not equal to the maximal number of test
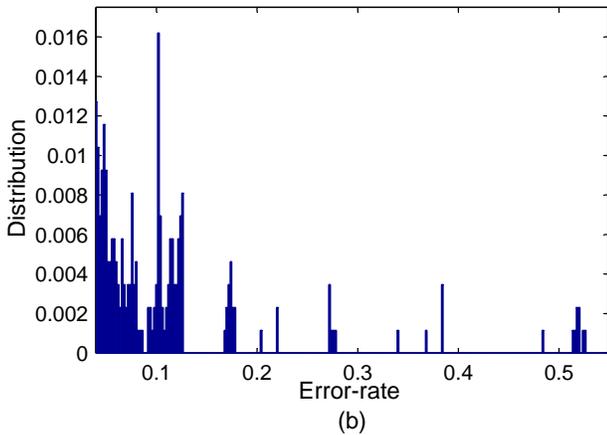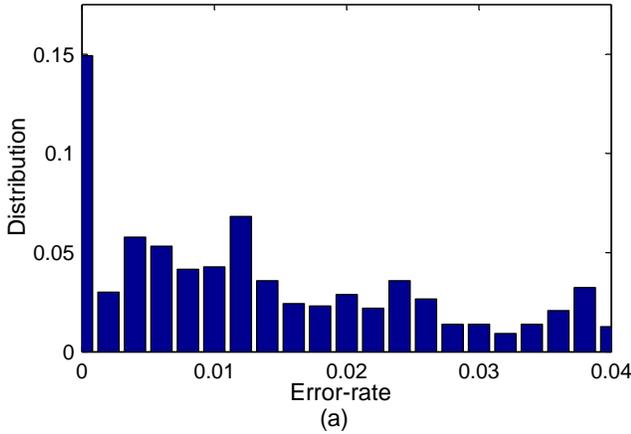
Fig.8. Normalized histogram of error-rates of 864 faulty circuits associated with C432_7. (a) The error-rate range is from 0 to 0.04. (b) The error-rate range is from 0.04 to 0.55. Note: the scales of (a) and (b) are different.
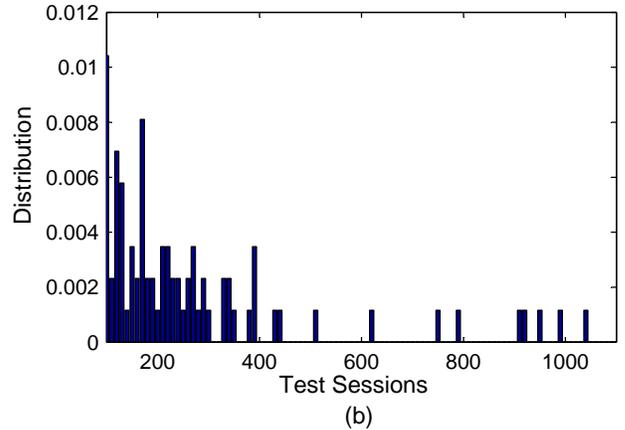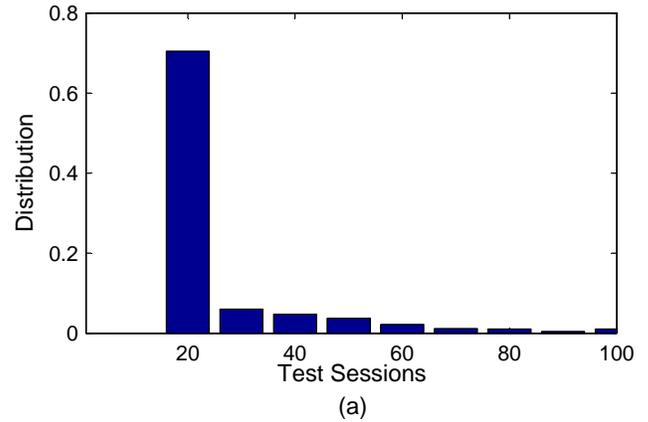
Fig.9. Normalized histogram of the number of test sessions applied when classifying the 864 faulty circuits of C432_7. (a) The number of test sessions is in the range of 1 to 100. (b) The number of test sessions is in the range of 100 to 1100. Note: the scales of (a) and (b) are different.

sessions 3182. This is because for some circuits, the estimated error-rate is either below 0.019 or above 0.02, and stopping condition is satisfied before running 3182 test sessions.

TABLE II
AVERAGE NUMBER OF TEST SESSIONS OF FAULTY C432_7 CIRCUITS FOR DIFFERENT ERROR-RATE RANGE

Error-rate Range	Average Number of Test Sessions	Number of Circuits in the Range
0~0.004	20	177
0.004~0.008	21.33	98
0.008~0.012	28.32	74
0.012~0.015	54.23	66
0.015~0.017	192.33	21
0.017~0.019	296.53	19
0.019~0.02	995.44	16
0.02~0.022	341.05	19
0.022~0.024	252.90	28
0.024~0.027	95.82	34
0.027~0.034	56.13	40
0.034~0.05	31.20	99
0.05~1	20.15	172

The percentage of circuits misclassified as type A is called *test escape*, and those misclassification as type B is called *yield loss*. Intuitively, if the error-rate of a circuit is far from the threshold, it is less likely to be misclassified. Because most of the 864 circuits in our experiment have an error-rate far from the threshold, the test escape and yield loss should be low. Table 3 lists the number of misclassification for three different error-rate ranges. Thus for this experiment the test escape is 10/471=3.33% and the yield loss is 13/393=2.12%.

The same classification experiment was applied to benchmark circuit C880. A single output circuit, namely C880_26, was derived from C880 such that one primary output of C880 is the only output of C880_26, and the other primary outputs of C880 are treated as internal wires of C880_26. Because C880_26 has 1960 single stuck-at faults, we obtained 1960 faulty copies of C880_26. Each faulty copy corresponds to a single stuck-at fault.

In the experiment, r_{th}=0.02, r_{thn}=0.019, β=0.05 and L=50. Table 4 lists the average number of test sessions for different sub-ranges of error-rate and the number of faulty circuits in each range. Similar to the results for C432_7, the number of test sessions increases when the error-rate range is close to 0.019~0.02. Table 4 also quantifies the number of misclassifications made. Of the 1554 type A circuits, 7 are misclassified. Of the 206 type B circuits, 6 are misclassified. The resulting yield loss is 0.45% and the test escape is 2.91%.

The results of these experiments on C432_7 and C880_26 are consistent with our analytical results. When the error-rate of a circuit is far from the threshold, correct classification

TABLE III
MISCLASSIFICATION IN THE EXPERIMENT FOR C432_7

Error-rate Range	Actual Type	Actual Number of Faulty Circuits	Number of Misclassified Circuits
0~0.019	Type A	455	1
0.019~0.02	Type A	16	9
0.02~1	Type B	393	13

decision can be made after a small number of test sessions. As the error-rate of a circuit approaches the threshold, more test sessions are needed. In our experiments, we choose β to be 0.05, which limits the probability of misclassification, and use β for the worse situation to determine other test parameters. When the error-rate of a circuit is far from the threshold, the probability of making a wrong decision is actually smaller than β. From table 4 it is seen that the probability of making a wrong decision increases as the actual error-rate approaches the threshold.

TABLE IV
EXPERIMENTAL RESULTS FOR FAULTY C880_26 CIRCUITS

Error-rate Range	Actual Type	Average Number of Test Sessions	Number of Actual Circuits in the Range	Number of Misclassified Circuits
0~0.004	A	20	1397	0
0.004~0.008	A	21.57	74	0
0.008~0.012	A	24.16	37	0
0.012~0.015	A	56.88	16	0
0.015~0.017	A	213.71	7	0
0.017~0.019	A	818	17	2
0.019~0.02	A	1657.83	6	5
0.02~0.022	B	873.76	17	4
0.022~0.024	B	244	7	0
0.024~0.027	B	80.38	8	1
0.027~0.034	B	68.14	7	1
0.034~0.05	B	34.97	38	0
0.05~1	B	20.19	129	0

In the experiment for C880_26, 1547 type A circuits are correctly classified as type A, and 6 type B circuits are misclassified as type A. Thus, 1553 circuits are classified as type A with 6 of them actually being type B. If we sell all 1553 circuits, in principle 6 might justifiable be returned. The fraction of chips that are beyond-tolerance (FBT), namely 6/1553, should not be confused with the classical notion of defects per million (DPM) used as a measure for product quality. We require that the maximal probability of misclassification for a circuit be β. Theoretically, for the worst scenario in classification, the number of actual type B circuits that are misclassified as type A is ($\beta \times$ the number of actual type B circuits); and the number of actually type A circuits that are misclassified as type B is ($\beta \times$ the number of actual type A circuits). Thus the number of circuits that are classified as type A is (the number of actual type A circuits $- \beta \times$ the number of actual type A circuits $+ \beta \times$ the number of actual type B circuits). Then, FBT is equal to

$$\frac{\beta \times the \ number \ of \ actual \ type \ B \ circuits}{the \ number \ of \ circuits \ that \ are \ classified \ as \ type \ A}.$$

In practice, the probability of misclassification is lower than β when the error-rate of a circuit is far from the threshold. Thus, the real value of FBT is usually lower then its theoretical value. In our experiment on the 1960 faulty circuits of C880_26, the experimental value of FBT is 6/1553(=0.38%), while the theoretical value of FBT is 0.05x206/(1554-0.05x1554+0.05x206) or 0.696%.

VIII. CONCLUSIONS

Error-tolerance mandates new domains for VLSI tests, such as error-rate. Here, the objective is to classify chips according to their error-rate rather than pass/fail, or *X* MHz vs. *Y* MHz. The fundamental task associated with this classification is the estimation of error-rate. An error-rate estimation technique based on signature analysis was presented in [11] and [12], where it was shown that this estimation and chip classification can be effectively implemented with reasonable BIST test resource. Another error-rate estimation technique based on ones counting also resulted in an estimator and classification technique [13]. The statistical properties of this estimator were not fully studied, some results were non-intuitive, and the required resources needed to carry out the test process appeared to be larger than necessary.

In this report, we again consider the problem addressed in [13], and develop a mathematically simpler estimator. In addition, we analyze the statistical characteristics of the estimator, such as its expectation, variance and probability density function. We determine the conditions when this estimator is biased vs. unbiased. Based on the statistical analysis, we describe a method for selecting the values for two key test parameters, namely the number of test patterns per session, *L*, and the number of test sessions, *S*. These parameters impact the test resources used for error-rate estimation. We show how these parameters are a function of both test time and storage requirements. In addition, we provide useful and practical bounds for these two parameters. The results of our analysis show that the test resources required for error-rate estimation based on ones counting are quite comparable to those needed when signature analysis is employed. Thus, the major difference in these two approaches is that signature based error-rate analysis can operate on a multi-output circuit, while ones counting processes one output at a time unless the test hardware is replicated for each output line.

We also address the problem of classifying chips based on their error-rate. The proposed classification procedure is partitioned into multiple phases. This process can significantly reduce classification time without any loss in the quality of classification.

Throughout this report several assumptions have been made, such as assuming a normal distribution for some random variables, and dropping terms that lead to second or third order effects. We have addressed these issues in our experiments, and validated that these assumptions are appropriate. We have also considered various boundary or extreme conditions, such as where the test length *L* of a test session is 1 or very large.

In our analysis, we assume all possible input patterns are equally likely to appear and have uniform distribution. Thus the test pattern generator in BIST is implemented with LFSR and generates pseudo random test patterns. For some circuits, their functional input may have different statistics than uniform distribution. To deal with this situation, test patterns can be obtained from functional test circuit instead of LFSR, and the same procedure is applied. Thus, the estimated error-rate is a conditional probability of outputting an error given the real statistics of input patterns. Thus, if the distribution of functional input patterns is known, it may be feasible to replace the PRPG with one that produces patterns whose distribution more closely mimics the real world.

For multiple output circuits, the output lines may not be of the same importance. Thus, when these output lines are considered independently, each of these lines will be associated with a different error-rate threshold. For example, the most significance bit line might have a much smaller error-rate threshold than the least significance bit line. Under this situation, the technique presented in this report can be used to estimate the error-rate of each output line and justify whether the error-rate is smaller than a threshold.

When the output lines of a multiple output circuit are considered as a whole for error-rate, its error-rate cannot be simply derived from the error-rate of each output line because of the correlation of errors among different output lines. This problem has been solved by using signature analysis [12], while using ones counting to solve this problem is the basis of a forthcoming paper.

APPENDIX

A. Expectation and Variance of the Estimator

For error-rate estimation based on ones counting compression, we propose using the estimator $\hat{r} = \dfrac{V_D}{L} + \dfrac{M_D{}^2}{L^2}$, where M_D and V_D are, respectively, the sample mean and sample variance of the sampled random variable *D*. In this appendix, we derive expressions for the expectation and variance of the estimator.

1) Preliminaries

In Section 3, we defined $D = \sum_{i=1}^{L} X_i$, where $X_1, X_2, \ldots, X_L$ are i.i.d. and have the same distribution as the random variable *X*. The PDF of *X* is Prob(*X*=1)=p_1, Prob(*X*= −1)=p_2, and Prob(*X*=0)=1−p_1−p_2, where $0 \le p_1 \le 1$, $0 \le p_2 \le 1$ and $0 \le p_1 + p_2 \le 1$.

To simplify future mathematical expressions, we define α_0, α_2, α_3 and α_4 as follows:

$$\alpha_0 = E\{X\} = p_1 - p_2,$$

$$\alpha_2 = E\{(X - \alpha_0)^2\} = p_1 + p_2 - (p_1 - p_2)^2 = p_1 + p_2 - \alpha_0^2,$$

$$\alpha_3 = E\{(X - \alpha_0)^3\} = E\{X^3 - 3X^2\alpha_0 + 3X\alpha_0^2 - \alpha_0^3\}$$

$$= (p_1 - p_2) - 3(p_1 + p_2)\alpha_0 + 3(p_1 - p_2)\alpha_0^2 - \alpha_0^3$$

$$= \alpha_0 - 3(p_1 + p_2)\alpha_0 + 2\alpha_0^3,$$

$$\alpha_4 = E\left\{(X - \alpha_0)^4\right\}$$
$$= E\left\{X^4 + 4X^2\alpha_0^2 + \alpha_0^4 - 4X^3\alpha_0 - 4X\alpha_0^3 + 2X^2\alpha_0^2\right\}$$
$$= (p_1 + p_2) + (6(p_1 + p_2) - 4)\alpha_0^2 - 3\alpha_0^4 .$$

α_0 is the expectation of X, and α_2, α_3 and α_4 are the 2nd, 3rd and 4th central moments of X.

Additional symbols are defined below.

$$\mu_0 = E\{D\} ,$$

$$Y = D - \mu_0 \text{ and } Y_i = D_i - \mu_0 \ (i = 1, 2, \ldots , S),$$

$$\mu_1 = E\{Y\} = E\{D - \mu_0\} = 0 ,$$

$$\mu_2 = E\{Y^2\} = E\{(D - \mu_0)^2\} ,$$

$$\mu_3 = E\{Y^3\} = E\{(D - \mu_0)^3\} ,$$

$$\mu_4 = E\{Y^4\} = E\{(D - \mu_0)^4\} .$$

μ_0 denotes the expectation of D. $Y_1, Y_2, \ldots , Y_S$ are i.i.d. and have the same distribution as Y. μ_1, μ_2, μ_3 and μ_4 are the 1st, 2nd, 3rd and 4th order moments of Y.

Since $D = \sum_{i=1}^{L} X_i$, we are able to express μ_0, μ_2, μ_3 and μ_4 in terms of α_0, α_2, α_3 and α_4.

$$\mu_0 = E\{D\} = LE\{X\} = L\alpha_0$$

$$\mu_2 = E\{Y^2\} = E\left\{(D - \mu_0)^2\right\} = Var\{D\}$$
$$= L(p_1 + p_2 - \alpha_0^2) = L\alpha_2$$

$$\mu_3 = E\{Y^3\} = E\left\{(D - \mu_0)^3\right\}$$
$$= E\left\{(X_1 + X_2 + \ldots + X_L - L\alpha_0)^3\right\}$$
$$= E\left\{\left(\sum_{i=1}^{L}(X_i - \alpha_0)\right)^3\right\} = LE\left\{(X - \alpha_0)^3\right\} = L\alpha_3$$

$$\mu_4 = E\{Y^4\} = E\left\{(D - \mu_0)^4\right\}$$
$$= E\left\{(X_1 + X_2 + \ldots + X_L - L\alpha_0)^4\right\}$$
$$= E\left\{\left(\sum_{i=1}^{L}(X_i - \alpha_0)\right)^4\right\}$$
$$= LE\left\{(X - \alpha_0)^4\right\} + 3L(L-1)\left(E\left\{(X - \alpha_0)^2\right\}\right)^2$$
$$= L\alpha_4 + 3L(L-1)\alpha_2^2$$

2) Derivation of the Expectation and Variance of the Estimator

The expectation of the sample variance of a random variable is equal to the expectation of the random variable [14]. So we have

$$E\{V_D\} = Var\{D\} . \tag{A1}$$

The expectation of M_D^2 is computed as follows.

$$E\{M_D^2\} = E\left\{\left(\frac{D_1 + D_2 + \ldots + D_S}{S}\right)^2\right\}$$
$$= \frac{1}{S^2}E\left\{\sum_{i=1}^{S}D_i^2 + \sum_{i=1}^{S}\sum_{j=1, j\neq i}^{S}D_iD_j\right\}$$
$$= \left(SE\{D^2\} + S(S-1)(E\{D\})^2\right)/S^2$$
$$= \left(E\{D^2\} - (E\{D\})^2 + S(E\{D\})^2\right)/S$$
$$= \left(Var\{D\} + S(E\{D\})^2\right)/S . \tag{A2}$$

Thus, the expectation of the estimator is

$$E\{\frac{V_D}{L} + \frac{M_D^2}{L^2}\} = \frac{1}{L}E\{V_D\} + \frac{1}{L^2}E\{M_D^2\}$$
$$= \frac{Var\{D\}}{L} + \frac{1}{SL^2}\left(Var\{D\} + S(E\{D\})^2\right)$$
$$= \mu_2/L + (\mu_2 + S\mu_0^2)/(SL^2)$$
$$= \alpha_2 + (L\alpha_2 + SL^2\alpha_0^2)/(SL^2)$$
$$= (p_1 + p_2) + [(p_1 + p_2) - (p_1 - p_2)^2]/SL . \tag{A3}$$

Next we compute the variance of the estimator.

$$Var\left\{\frac{V_D}{L} + \frac{M_D^2}{L^2}\right\} = Var\left\{\frac{V_D}{L}\right\} + Var\left\{\frac{M_D^2}{L^2}\right\} + 2Cov\left(\frac{V_D}{L}, \frac{M_D^2}{L^2}\right)$$
$$= \frac{1}{L^2}Var\{V_D\} + \frac{1}{L^4}Var\{M_D^2\} + \frac{2}{L^3}Cov(V_D, M_D^2). \tag{A4}$$

We next compute each component in (A4). Now

$$Var\{M_D^2\} = E\{M_D^4\} - \left(E\{M_D^2\}\right)^2 \tag{A5}$$

where

$$E\{M_D^2\} = \frac{1}{S}Var\{D\} + (E\{D\})^2 = \frac{\mu_2}{S} + \mu_0^2$$

and

$$E\{M_D^4\} = E\left\{\left(\frac{D_1 + D_2 + \ldots + D_S}{S}\right)^4\right\}$$
$$= E\left\{\left(\frac{Y_1 + Y_2 + \ldots + Y_S}{S} + \mu_0\right)^4\right\}$$
$$= E\left\{\left(\sum_{i=1}^{S}Y_i/S\right)^4 + 4\mu_0\left(\sum_{i=1}^{S}Y_i/S\right)^3 + 6\mu_0^2\left(\sum_{i=1}^{S}Y_i/S\right)^2 + 4\mu_0^3\left(\sum_{i=1}^{S}Y_i/S\right) + \mu_0^4\right\}$$
$$= \left(S\mu_4 + 3S(S-1)\mu_2^2\right)/S^4 + 4\mu_0\left(S\mu_3\right)/S^3 + 6\mu_0^2\left(S\mu_2\right)/S^2 + \mu_0^4$$
$$= \frac{\mu_4}{S^3} + \frac{3(S-1)\mu_2^2}{S^3} + \frac{4\mu_0\mu_3}{S^2} + \frac{6\mu_2\mu_0^2}{S} + \mu_0^4 . \tag{A6}$$

Thus,

$$Var\{M_D^2\} = \frac{\mu_4}{S^3} + \frac{3(S-1)\mu_2^2}{S^3} + \frac{4\mu_0\mu_3}{S^2} + \frac{6\mu_2\mu_0^2}{S} + \mu_0^4 - \left(\frac{\mu_2}{S} + \mu_0^2\right)^2$$
$$= \frac{\mu_4 - 3\mu_2^2}{S^3} + \frac{2\mu_2^2 + 4\mu_0\mu_3}{S^2} + \frac{4\mu_2\mu_0^2}{S} . \tag{A7}$$

We next consider the term $Var\{V_D\}$. Let $M_Y = (Y_1 + Y_2 + \ldots + Y_4)/S = M_D - \mu_0$. Then

$$V_D = \frac{1}{S-1}\sum_{i=1}^{S}(D_i - M_D)^2$$
$$= \frac{1}{S-1}\sum_{i=1}^{S}((D_i - \mu_0) - (M_D - \mu_0))^2$$

$$= \frac{1}{S-1} \sum_{i=1}^{S} (Y_i - M_Y)^2 \cdot$$

Recall that

$$Var\{V_D\} = E\{V_D^2\} - (E\{V_D\})^2$$

$$= E\{V_D^2\} - (Var\{D\})^2 = E\{V_D^2\} - \mu_2^2 . \tag{A8}$$

Hence,

$$E\{V_D^2\} = E\left\{ \left(\sum_{i=1}^{S} (Y_i - M_Y)^2 \Big/ (S-1) \right)^2 \right\}$$

$$= \frac{1}{(S-1)^2} E\left\{ \left(\sum_{i=1}^{S} Y_i^2 - SM_Y^2 \right)^2 \right\}$$

$$= \frac{1}{(S-1)^2} \left(E\left\{ \left(\sum_{i=1}^{S} Y_i^2 \right)^2 \right\} - 2S^2 E\{Y_1^2 M_Y^2\} + S^2 E\{M_Y^4\} \right) . \tag{A9}$$

In (A9),

$$E\left\{ \left(\sum_{i=1}^{S} Y_i^2 \right)^2 \right\} = E\{SY_1^4 + S(S-1)Y_1^2 Y_2^2\}$$

$$= S\mu_4 + S(S-1)\mu_2^2 \tag{A10}$$

and

$$E\{Y_1^2 M_Y^2\} = E\{Y_1^2 (Y_1 + Y_2 + ... + Y_S)^2 / S^2\}$$

$$= (\mu_4 + (S-1)\mu_2^2) / S^2 \tag{A11}$$

and

$$E\{M_Y^4\} = E\{(Y_1 + Y_2 + ... + Y_S)^4 / S^4\}$$

$$= (\mu_4 + 3(S-1)\mu_2^2) / S^3 . \tag{A12}$$

Therefore,

$$E\{V_D^2\} = \frac{\mu_4}{S} + \frac{S^2 - 2S + 3}{S(S-1)} \mu_2^2 . \tag{A13}$$

So finally for (A8) we have

$$Var\{V_D\} = E\{V_D^2\} - \mu_2^2 = \frac{\mu_4}{S} - \frac{\mu_2^2}{S-1} + \frac{3\mu_2^2}{S(S-1)} . \tag{A14}$$

We now consider the term $Cov(V_D, M_D^2) = E\{V_D M_D^2\} - E\{V_D\}E\{M_D^2\}$ in (A4), where

$$E\{V_D M_D^2\} = E\left\{ \frac{1}{S-1} \sum_{i=1}^{S} (D_i - M_D)^2 M_D^2 \right\}$$

$$= E\left\{ \frac{1}{S-1} \sum_{i=1}^{S} (Y_i - M_Y)^2 M_D^2 \right\}$$

$$= \frac{1}{S-1} E\left\{ \left(\sum_{i=1}^{S} Y_i^2 - SM_Y^2 \right) M_D^2 \right\}$$

$$= \frac{\mu_4 - 2\mu_0\mu_3}{S(S-1)} + \frac{\mu_2^2}{S} + \frac{2\mu_0\mu_3}{S-1} + \mu_2\mu_0^2 - \frac{\mu_4}{S^2(S-1)} - \frac{3\mu_2^2}{S^2} \tag{A15}$$

and

$$E\{V_D\}E\{M_D^2\} = Var\{D\}E\{M_D^2\}$$

$$= \mu_2 \left(\frac{\mu_2}{S} + \mu_0^2 \right) = \frac{\mu_2^2}{S} + \mu_2\mu_0^2 . \tag{A16}$$

Therefore,

$$Cov(V_D, M_D^2) = \frac{\mu_4 - 2\mu_0\mu_3}{S(S-1)} + \frac{2\mu_0\mu_3}{S-1} - \frac{\mu_4}{S^2(S-1)} - \frac{3\mu_2^2}{S^2} . \tag{A17}$$

Substituting $Var\{M_D^2\}$, $Var\{V_D\}$ and $Cov(V_D, M_D^2)$ into (A4) using (A7), (A14) and (A17), the variance of the estimator is obtained as

$$Var\left\{ \frac{V_D}{L} + \frac{M_D^2}{L^2} \right\} = \frac{1}{L^2} \left(\frac{\mu_4}{S} - \frac{\mu_2^2}{S-1} + \frac{3\mu_2^2}{S(S-1)} \right)$$

$$+ \frac{1}{L^4} \left(\frac{\mu_4 - 3\mu_2^2}{S^3} + \frac{2\mu_2^2 + 4\mu_0\mu_3}{S^2} + \frac{4\mu_2\mu_0^2}{S} \right)$$

$$+ \frac{2}{L^3} \left(\frac{\mu_4 - 2\mu_0\mu_3}{S(S-1)} + \frac{2\mu_0\mu_3}{S-1} - \frac{\mu_4}{S^2(S-1)} - \frac{3\mu_2^2}{S^2} \right) . \tag{A18}$$

Because S is large compared to 1, we assume that $S-1 \approx S$. Now we have

$$Var\left\{ \frac{V_D}{L} + \frac{M_D^2}{L^2} \right\} = \frac{1}{L^2} \left(\frac{\mu_4 - \mu_2^2}{S} + \frac{3\mu_2^2}{S^2} \right)$$

$$+ \frac{1}{L^4} \left(\frac{\mu_4 - 3\mu_2^2}{S^3} + \frac{2\mu_2^2 + 4\mu_0\mu_3}{S^2} + \frac{4\mu_2\mu_0^2}{S} \right)$$

$$+ \frac{2}{L^3} \left(\frac{\mu_4 - 2\mu_0\mu_3 - 3\mu_2^2}{S^2} + \frac{2\mu_0\mu_3}{S} - \frac{\mu_4}{S^3} \right)$$

$$= \frac{2\alpha_2^2}{S} + \frac{3\alpha_2^2}{S^2} + \frac{\alpha_4 - 3\alpha_2^2 + 4\alpha_2\alpha_0^2 + 2\alpha_0\alpha_3}{LS} + \frac{2\alpha_0\alpha_3}{LS^2}$$

$$+ \frac{4\alpha_0\alpha_3 + \alpha_4 - \alpha_2^2}{L^2 S^2} - \frac{3\alpha_2^2}{LS^3} + \frac{\alpha_4 - 3\alpha_2^2}{L^2 S^3} + \frac{\alpha_4 - 3\alpha_2^2}{L^3 S^3} . \tag{A19}$$

B. Q Function: Q(x)

Q function $Q(x)$ computes the right tail probability of normal distribution $N(0, 1)$. So Q function is defined as

$$Q(x) = \int_{x}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt .$$

Using Q function, many probability related to a random variable having normal distribution $N(\mu, \sigma^2)$ can be simply expressed simply. Here are some examples. Let T denote such a random variable.

Probability of T being in $(x, +\infty)$ is

$$Q((x-\mu)/\sigma) = \int_{x}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t-\mu)^2/2\sigma^2} dt .$$

Probability of T being in $(-\infty, x)$ is $1 - Q((x-\mu)/\sigma)$.

Probability of T being in $(\mu-\alpha, \mu+\alpha)$, where $\alpha > 0$, is $1 - 2Q(\alpha/\sigma)$.

C. Justification of (25) with r<1/5 being Maximized when $\alpha_0 = 0$ in Section 4.4

Equation (25) in Section 4.4 states that

$$S = \frac{(Q^{-1}((1-\gamma)/2))^2}{r^2\varepsilon^2} \left(2(r - \alpha_0^2)^2 + \frac{(r - 3r^2) + (10r - 2)\alpha_0^2 - 6\alpha_0^4}{L} \right).$$

Because $\alpha_0 = p_1 - p_2$ and $r = p_1 + p_2$, we have $0 \le \alpha_0^2 < r$ and $0 < r - \alpha_0^2 \le r$. Thus $(r - \alpha_0^2)$ is maximized when $\alpha_0 = 0$. So is $(r - \alpha_0^2)^2$. The term $T = (r - 3r^2) + (10r - 2)\alpha_0^2 - 6\alpha_0^4$ can be rewritten as

$$T = -6\left(a_0^2 - (5r-1)/6\right)^2 + (7r^2 - 4r + 1)/6. \tag{C1}$$

When $\alpha_0^2 > (5r-1)/6$, T decreases as α_0^2 increases. For r <1/5, $(5r-1)/6$<0. On the other hand, α_0^2 is always non-negative. For r <1/5, T is maximized when $\alpha_0^2 = 0$. Thus, for r<1/5, both $(r - \alpha_0^2)^2$ and $-6\left(a_0^2 - (5r-1)/6\right)^2 + (7r^2 - 4r + 1)/6$ are maximized when $\alpha_0^2 = 0$. Hence, for r <1/5,

$$S = \frac{\left(Q^{-1}((1-\gamma)/2)\right)^2}{r^2 \varepsilon^2}\left(2\left(r - \alpha_0^2\right)^2 + \frac{(r - 3r^2) + (10r - 2)\alpha_0^2 - 6\alpha_0^4}{L}\right)$$

is maximized when $\alpha_0^2 = 0$, and its maximum is

$$S = \frac{\left(Q^{-1}((1-\gamma)/2)\right)^2}{\varepsilon^2}\left(2 + \frac{1}{L}\left(\frac{1}{r} - 3\right)\right). \tag{C2}$$

REFERENCES

[1] International Technology Roadmap for Semiconductors: Yield Enhancement, 2006 Update, http://www.itrs.net/Links/2006Update/2006UpdateFinal.htm.

[2] M. Butts, A. DeHon and S. C. Goldstein, "Molecular Electronics: Devices, Systems and Tools for Gigagate, Gigabit Chips", *Proc. Int'l. Conf. on Computer Aided Design*, pp. 433-440, Nov. 10-14, 2002.

[3] R. I. Bahar, "Trends and Future Directions in Nano Structure Based Computing and Fabrication", *Proc. 24th Int'l. Conf. on Computer Design*, Oct. 1-4, 2006.

[4] I. Koren and Z. Koren, "Defect Tolerance in VLSI Circuits: Techniques and Yield Analysis", *Proc. IEEE*, Vol. 86, No. 9, pp. 1819-1838, Sept. 1998.

[5] M. A. Breuer, S. K. Gupta and T. M. Mak, "Defect and Error Tolerance in the Presence of Massive Numbers of Defects", *IEEE Design and Test of Computers*, pp. 216-227, May-June 2004.

[6] M. A. Breuer, "Intelligible Test Techniques to Support Error-Tolerance", *Proc. 13th Asian Test Symp.*, pp. 386-393, 2004.

[7] H. H. Kuok, *Audio Recording Apparatus Using an Imperfect Memory Circuit*, US Patent 5414758, Patent and Trademark Office, 1995; http://www.freepatentsonline.com/5414758.html.

[8] P. V. Argade, *Digital Secretary*, US patent 5651055, Patent and Trademark Office, 1997; http://freepatentsonline.com/5651055.html.

[9] H. Chung and A. Ortega, "Analysis and Testing for Error Tolerant Motion Estimation", *Proc. IEEE Int'l. Symp. on Defect and Fault Tolerance in VLSI Systems*, pp. 514-522, Nov. 2005.

[10] H. Chung and A. Ortega, "System Level Fault Tolerance for Motion Estimation", Technical Report USC-SIPI354, Signal and Image Processing Institute, Univ. of Southern California, 2002.

[11] M. A. Breuer, "Estimating Error Rate in Error Tolerant VLSI Chips", *IEEE Int'l Workshop on Electronic Design, Test and Applications*, pp. 321-326, January 2004.

[12] Z. Pan and M. A. Breuer, "Estimating Error Rate in Defective Logic Using Signature Analysis", *IEEE Trans. on Computers*, Vol. 56, No. 5, pp. 650-661, May 2007.

[13] S. Shahidi and S. K. Gupta, "Estimating Error Rate During Self-Test via One's Counting", *Proc. Int'l. Test Conf.*, paper 15.3, 2006.

[14] G. Casella and R. L. Berger, *Statistical Inference*, Duxbury Press, 2nd ed., 2001.

[15] M. A. Breuer and H. Zhu, "Error-Tolerance and Multi-Media," *Proc. Int'l. Conf. on Intelligent Info. Hiding and Multimedia*, pp. 521-524, 2006.

[16] M. Abramovici, M. A. Breuer and A. D. Friedman, *Digital Systems Testing and Testable Design*, Wiley-IEEE Press, 1994.

[17] N. K. Jha and S. K. Gupta, *Testing of Digital Systems*, Cambridge Univ. Press, 1st ed., 2002.