University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

CSE Conference and Workshop Papers

Computer Science and Engineering, Department
of

1999

Empirical Computation of Reject Ratio in VLSI Testing

Shashank K. Mehta

University of Nebraska-Lincoln, smehta@cse.unl.edu

Sharad C. Seth

University of Nebraska-Lincoln, seth@cse.unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/cseconfwork>

Part of the [Computer Sciences Commons](#)

Mehta, Shashank K. and Seth, Sharad C., "Empirical Computation of Reject Ratio in VLSI Testing" (1999).
CSE Conference and Workshop Papers. 22.

<https://digitalcommons.unl.edu/cseconfwork/22>

This Article is brought to you for free and open access by the Computer Science and Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in CSE Conference and Workshop Papers by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Empirical Computation of Reject Ratio in VLSI Testing *

Shashank K. Mehta [†] Sharad C. Seth
Dept. of Computer Science and Engineering
University of Nebraska-Lincoln
Lincoln, NE 68588-0115
smehta, seth @cse.unl.edu

Abstract

Among significant components of testing cost are test-length, reject ratio, and lost-yield. In this paper a new approach is proposed to estimate the reject ratio. The empirical model is based on test-data properties that are believed to be invariant for a wide range of manufacturing technologies and types of tests. The analysis is carried out entirely in terms of the device test data, as might be available from wafer probe. Experimental results demonstrate robustness of the model.

1. Introduction

Testing accounts for substantial costs in the manufacture of semiconductor devices (“chips”). According to industrial sources, the already substantial testing costs are expected to rise further in the next decade [6]. Test related costs are incurred at several points during the entire design and manufacturing cycle. A recent cost model [11] identifies four major cost components of testing related to test preparation, test execution, test-related silicon costs, and imperfect test quality. The last component is the focus of this paper. We propose an analytical approach for accurately estimating the ‘reject ratio’ [1] (also called, ‘escape rate’ or ‘defect level’ in the literature). The results may also help ascertain yield loss due to non-functional testing methods, such as IDDQ and scan.

Test plans for currently manufactured devices involve some combination of tests for functional, scan, IDDQ, delay, and possibly, bridging faults. As no single type of test is believed to catch all defective parts, cost-benefit analysis is made for including different types of tests and deciding

their order of application. Given that test preparation and application costs generally rise with test length, the question arises: when is the point of diminishing return reached for a particular type of test? It has been shown that the stuck-at fault coverage of a test is a very unreliable measure to answer this question [4]. In this paper we provide an alternative way of answering the question.

1.1. Past Work

Recognizing that the reject ratio, while easy to define, is hard to estimate through direct measurement, researchers have proposed several alternative approaches for this purpose during the past twenty years. Common to all the methods is statistical modeling of the circuit testability, the chip testing process, or both. However, the earlier models [1, 3, 7, 10, 12] fail to account adequately for the frequent jumps (discontinuities) in the chip tester data.

In earlier work [2], we posited such jumps to be an essential characteristic of the tester data and accounted for them in a statistical model by introduction of fault *latency* as a new parameter. The model provided an expression for the *yield* of chips after application of n test vectors in terms of fault latencies and detectabilities as parameters. These parameters are estimated directly from the tester data, as is the process yield, thus allowing an estimate of the field reject ratio.

An attractive feature of the above approach is its independence of any specific fault model and fault coverage. These are implicitly accounted for by the latency and detectability parameters. Because these parameters are estimated from the tester data, the model can self-adjust for any inadequacies of the test.

Our recent investigations indicate that the reject ratio estimates by the latency based approach may be unduly optimistic and not quite robust. One source of error can be traced to the way the process yield is estimated from the tester data; even a slight error in the estimate can have a sig-

*This work was supported by NSF, Grant MIP-9806795 and by UNL Center for Communication and Information Science.

[†] Home institute: University of Pune, Pune, India

nificant effect on the calculated reject ratio. Another source of error comes from the assumption that the range of fault latencies is unbounded but all accounted for by the end of the testing process.

Our proposed approach differs from the previous ones in that no explicit failure mechanism is assumed in our model. In the prior studies, the modeling parameters came from such assumptions, e.g., fault coverage and yield [10, 12]; fault coverage, yield, and the average number of faults on a defective chips [1]; Stapper's two yield parameters and the average number of faults per defect [7]; and yield, fault latencies, and fault-detection probabilities [2]. We believe the failure mechanisms to be too numerous and complex to be captured accurately by these means. Instead, we derive our model parameters from commonly observed characteristics of faults, testing process, and test data that appear to be invariant across a range of manufacturing processes.

An advantage of the proposed approach is that a yield estimate is not required, thus, it can be applied also during the yield-learning phase of manufacturing. All previous methods require a yield estimate and their results are very sensitive to this value. On the other hand, because the approach requires tester data for analysis, an obvious disadvantage is that the proposed analysis can only be carried out after a batch of devices have been fabricated and tested.

2. An Empirical Model for Testing

Our method is based on two statistical quantities, namely event-probability and event-size distribution probability. We believe that these probabilities are well-behaved functions of test vectors in every test result and can be modeled by simple functions. These functions are used to model the cumulative fallout (number of detected faulty devices). The estimate for the reject ratio is determined by computing the asymptotic value of the cumulative fallout.

In the rest of this section we first state and justify the general assumptions used in our model regarding the faults, the testing process, and the test data. As illustration, we use wafer test data for a high-volume digital CMOS device obtained from Delco Electronics; the wafer test for the device covered 99.7% of the stuck-at faults [3]. Future work will include validation of the model in terms of other test data sets available to us.

2.1. Fault Characteristics

We distinguish between faults or functional deviance from physical defects (or simply, defects). The latter occur during device fabrication and may cause faults. Unfortunately, it is hard to characterize precisely which defects are likely to occur and with what frequency. Further, the commonly used fault models capture only a small subset of

all faults and the relationship between defects and modeled faults is not always easy to establish. In view of these uncertainties, we state and justify two rather general assumptions about defects and faults that are useful in analyzing device test data.

Assumption 1 Device defects cause varying degree of damage from the testing viewpoint, i.e., *some defects are detectable by a very large number of test vectors while others are detectable by only a few vectors.*

The assumption can be justified in terms of defect size which varies over a wide range, causing varying degree of damage in the logical domain, see [8]. Also, the vulnerability of the chip function varies with the location of the damage. Further justification for the assumption may be found in the great variance observed, typically, in the detectability of stuck-at faults in a circuit.

Assumption 2 *Defects that affect many devices occur less frequently than those that affect fewer devices.*

This assumption is true as a consequence of the yield learning that takes place with maturation of a fabrication process. In order to improve the yield it is most beneficial to eliminate defects in order of frequency of occurrence. Thus effort would be directed towards identifying and removing the causes of frequently-occurring defects.

2.2. Testing-Process Characteristics

Test vectors may be functional, generated randomly, or generated by targeting faults based on a fault model. In this section we argue that even when a fault model is used, the test generator has only a partial information and the characteristics of the test results do not differ, especially towards the end, no matter how the test vectors are generated.

As already discussed, fault models have a rather uncertain relation with the actual defects that are themselves hard to characterize precisely. The most significant information which is not available to a test generator is the number of chips affected by a given fault (the fault occurrence probability). Another unknown is which set of faults are caused by the same defect. Ideally a test generator needs to generate test for only one of them. Aside from this, most fault models tend to ignore complex faults. It is often difficult to generate tests without assumptions such as single-fault, no-bridging-fault, etc.

While targeting one fault, several additional faults are also detected. It is justified to assume that the chances of detection of these additional faults should be about the same even if the test vectors were generated randomly. As the test progresses, the set of undetected faults begins to be dominated more and more by the unmodeled faults. Since unmodeled or uncovered faults can only be detected by happenstance, in the later part of a test the fallout should be materially independent of how the test was generated.

Figure 1. Fallout vs. vector number

The defect level (reject ratio) of the final yield depends only on these uncovered or unmodeled faults so it should be possible to estimate it by extrapolating a function that models the later results with high fidelity.

2.3. Test-Data Characteristics

Figure 1 shows typical per-vector device fallout due to a test. Figure 2 shows the cumulative fallout. It is clear from the figures that there is a large variance in the fallout size. In this case a natural question arises, namely, can there exist a smooth (differentiable) curve which could be considered a good approximation to the cumulative fallout.

In the initial part of the test the variance in the fallout is so large that no smooth curve can possibly be considered a good fit. But in the later part the fallout size varies in a narrow range so a good fit could be found. This assertion was verified for our test data set using standard measures of goodness-of-fit [9].

Since the reject ratio depends on the number of chips that would have been found faulty had the test been continued, it could be estimated with a significant certainty by fitting a smooth curve on the the cumulative fallout curve and extrapolating it.

2.4. Structure of the Model

As discussed above, we shall view testing as a random process. $f(t)$ will denote the random variable representing the number of chips found faulty on the application of test vector V_t for $1 \leq t \leq T$. N will denote total number of chips.

Modeling the testing process involves development of a parametric expression, $\bar{f}_M(t)$, approximating the expectation value of $f(t)$.

Figure 2. Cumulative fallout

Reject Ratio Computation

In order to estimate the reject ratio, first, the parameter values should be determined so that $\bar{f}_M(t)$ fits well to the actual instance of $f(t)$. Then $N - \sum_{t'=1}^t \bar{f}_M(t')$ and $\sum_{t'=t+1}^{\infty} \bar{f}_M(t')$ respectively approximate the total number of chips and the number of bad chips that survived the first t test vectors. Thus the reject ratio after t is approximated by

$$r(\bar{f}_M, t) = \frac{\sum_{t'=t+1}^{\infty} \bar{f}_M(t')}{N - \sum_{t'=1}^t \bar{f}_M(t')} \quad (1)$$

The expression can be rewritten as

$$r(\bar{f}_M, t) = r_T(\bar{f}_M, t) + \frac{\sum_{t'=T+1}^{\infty} \bar{f}_M(t')}{N - \sum_{t'=1}^t \bar{f}_M(t')}$$

where r_T denotes the contribution to the reject ratio by the faulty chips detected by the remaining test vectors, i.e., $v_{t'}$ for $t < t' \leq T$.

$r_T(\bar{f}_M, t)$ is useful in checking the robustness of the model by comparing it with $r_T(f, t)$, which can be computed from the actual data. Table 1 shows the values of $r_T(f, t)$ and $r_T(\bar{f}_M, t)$ for various values of t in our experiment with Delco data.

Expectation Value of $f(t)$

Definition We consider failure of one or more chips at vector V_t (i.e., $f(t) > 0$) as an *event at t* . Let a binary random variable $\epsilon(t)$ characterize the event at t . Its value 1 shall denote that event has occurred at t with *size* $f(t)$.

The expectation value of $f(t)$ is given by

$$\begin{aligned} \bar{f}(t) &= \sum_{\sigma=0}^{m(t)} \sigma \cdot \text{Prob}(f(t) = \sigma) \\ &= \sum_{\sigma=1}^{m(t)} \sigma \cdot \text{Prob}(f(t) = \sigma \mid \epsilon(t) = 1) \cdot \text{Prob}(\epsilon(t) = 1) \end{aligned} \quad (2)$$

Figure 3. Cumulative events vs vector number

where $Prob(f(t) = \sigma | \epsilon(t) = 1)$, *event-size distribution probability*, denotes the probability of V_t detecting σ faulty chips given that event occurs at time t and $Prob(\epsilon(t) = 1)$, *event probability*, denotes the probability of an event at time t . It is assumed that event size is bounded above by $m(t)$.

Event Probability

Assumption 1 suggests that the defects with higher *visibility* (number of test vectors that can detect it) are likely to be detected earlier in the test. Therefore it is expected that the event probability should decrease as the test proceeds. Figure 3 shows the cumulative number of events, $\sum_{t'=1}^t \epsilon(t')$. Here we develop a model based on a simpler assumption: the probability of an event is proportional to the number of undetected faults.

Let $F(t)$ denote the number of faults remaining undetected after t test vectors. If we assume that the rate of detection of new faults is proportional to the number of undetected faults, then $F(t)$ can be approximated by $F(0)e^{-c \cdot t}$.

Assume that every input vector detects a set of S faults, where S is a random variable with a narrow range. Of the total possible subsets of S faults, $\binom{F(0)}{S}$, the number of subsets that only contain the faults that are already detected by the first $t-1$ vectors is $\binom{F(0) - F(t-1)}{S}$. Therefore, on the basis of our assumption that test vectors are generated randomly, the probability of no event (i.e., V_t does not detect any new fault) is $\binom{F(0) - F(t-1)}{S} / \binom{F(0)}{S}$. Assuming that S is small, it approximates to $(1 - e^{-c \cdot t})^S$. Denoting the average value of S by d , probability of event is given by

$$Prob(\epsilon(t) = 1) = 1 - (1 - e^{-c \cdot t})^d \quad (3)$$

Figure 4. Cumulative events with model

Figure 4 shows the best fit of the event curve by Equation 3. The fit is especially good in the second half of the test.

Event Size Distribution

From Assumption 2 it can be deduced that the faults that affect large number of chips, occur less frequently. On this basis we conclude that event-size distribution probability should have negative correlation with σ .

In this study, a simple model for the event size distribution is used, namely, $Prob(f(t) = \sigma | \epsilon(t) = 1) \propto 1/\sigma^k$ for $1 \leq \sigma \leq m(t)$, where $m(t)$ denotes the size of largest possible fallout by V_t .

Figure 5 shows the event size distributions in various windows of 1000 test vectors and their approximation by the model curves $1/\sigma^k$. The optimum values of k in each case is found to vary slightly so we replace k by

$$k(t) = a + b \cdot t \quad (4)$$

Normalizing the function, we get the expression for event-size probability

$$Prob(f(t) = \sigma | \epsilon(t) = 1) = (1/\sigma^{k(t)}) \cdot (1 / \sum_{\sigma=1}^{m(t)} (1/\sigma^{k(t)})), 1 \leq \sigma \leq m(t) \quad (5)$$

As observed earlier in this section, the variance of the fallout, $m(t)$, decreases quickly as the test progresses. In this work we have modeled $m(t)$ in the same way as the event probability so

$$m(t) = m_0 \cdot (1 - (1 - e^{-g \cdot t})^h) \quad (6)$$

Combining the Parts

Substituting Equations 3 and 5 into Equation 2 and approximating discrete sum by integral of the corresponding continuous curve, average fallout at time t is

Figure 5. Event size distribution in windows of 1000 vectors, 1-1000, 1001-2000, 2001-3000 (top row), 3001-4000, 4001-5000, and 5001-6000 (bottom row)

$$\bar{f}_M(t) = \frac{(1 - (1 - e^{-c \cdot t})^d) \cdot (k(t) - 1) \cdot (m(t)^{2-k(t)} - 1)}{(2 - k(t)) \cdot (1 - m(t)^{1-k(t)})} \quad (7)$$

where $k(t)$ and $m(t)$ are given by Equations 4 and 6 respectively.

3. Model Verification and Results

In this section we test the robustness of the model, Equation 7, by comparing the actual and the predicted reject ratios.

Using the values of c and d computed independently by fitting the cumulative event model to its data, a set of values of the parameters a, b, m_0, g, h was determined by fitting the curve $\sum_t \bar{f}_M(t)$ on the cumulative fallout curve of Figure 2 ignoring the test data for the first 2000 vectors. The data of the first 2000 vectors was not considered because of the enormous variance in this region which has undesirable effect on the accuracy of the fit in the remaining curve.

Figure 6 shows the actual and model cumulative-fallout curve. Since the model curve represents the expected value of the fallout, it can be considered a very good fit to the actual data which are an instance of the fallout.

To determine the goodness of the fit, partial reject ratios, r_T , were computed from the actual data and the model. These ratios, as parts per million, are given in Table 1. The reject ratios computed based on the model show close correspondence with those computed from the data. This indi-

Figure 6. Cumulative fallout with model

cates that the model should accurately determine the order of the actual reject ratio, i.e., $r_{model} = \Theta(r_{actual})$.

3.1. Prediction

The reject ratio was estimated using Equation 1. Limiting value of the cumulative fallout, the numerator of r.h.s. of the equation, was estimated by extrapolating the model curve of Figure 6. The value of the estimated reject ratio for various parameter values, determined from different local optima, varied from 1305 ppm to 2480 ppm for this data set of Delco test set. The narrow range of variance indicates that the model is robust.

t	2000	3000	4000	5000	6000	7000	8000	9000	10000	11000	12000
Data	193162	112451	104441	99907	89488	53522	50762	20839	15933	11673	1957
Model	193583	157005	126759	101138	79202	60379	44285	30634	19182	9704	1975

Table 1. Partial reject ratio, r_T , in parts per million

4. Conclusion

The proposed approach allows, by extrapolation of the test results, to estimate how many more devices would eventually fail if the test were to be carried out indefinitely. It needs to be noted that the criterion used for determining what is a good vs bad device may differ with test type. For example, scan tests may reject a functional device that fails on a non-functional vector. Similarly, delay test may reject a device that is functional but does not meet the delay requirement. Thus, the extrapolation would give a value of the reject ratio that is specific to a type of test.

Recent pilot studies have produced test results for different types of tests (scan, functional, etc.) on the same set of devices [4, 5]. With these data sets it will be possible to study and compare the projections of our model for the various types of tests. We are also considering how the model can be extended to estimate reject ratio when a combination of tests is applied to a batch of devices. There is enough information available in the data sets to verify any extensions to combined tests.

Acknowledgment The authors wish to thank Mr. Hailong Cui for timely help in producing the computational results and figures reported here.

References

- [1] V. D. Agrawal, S. C. Seth, and P. Agrawal. Fault coverage requirements in production testing of lsi circuits. *IEEE Journal of Solid State Circuits*, SC-17:57–61, February 1982.
- [2] D. V. Das, S. C. Seth, and V. D. Agrawal. Accurate computation of field reject ratio based on fault latency. *IEEE Trans VLSI Systems*, pages 537–545, December 1993.
- [3] D. V. Das, S. C. Seth, P. T. Wagner, J. C. Anderson, and V. D. Agrawal. An experimental study on reject ratio prediction for vlsi circuits: Kokomo revisited. *Proceedings of the International Test Conference*, pages 712–720, September 1990.
- [4] P. C. Maxwell, R. C. Aitken, V. Johansen, and I. Chiang. The effect of different test sets on quality level prediction: When is 80% better than 90%? *Proc. International Test Conference*, pages 358–364, 1991.
- [5] P. Nigh, W. Needham, K. Butler, and R. Aitken. An experimental study comparing the relative effectiveness of func-

- tional, scan, iddq, and delay-fault testing. *Proc. IEEE VLSI Test Symposium*, pages 459–454, 1997.
- [6] Sematech, Inc. The national technology roadmap for semiconductors. <http://www.sematech.org/public/roadmap/index.html>.
- [7] S. C. Seth and V. D. Agrawal. Characterizing the lsi yield equation from wafer test data. *IEEE Transactions on Computer Aided Design*, CAD-3:123–126, April 1984.
- [8] C. H. Stapper. Integrated circuit yield management and yield analysis: Development and implementation. *IEEE Trans Semiconductor Manufacturing*, pages 95–102, May 1995.
- [9] K. S. Trivedi. *Probability & Statistics with Reliability, Queuing, and Computer Science Applications*. Prentice Hall, Englewood Cliffs, NJ, 1982.
- [10] R. L. Wadsack. Fault coverage in digital integrated circuits. *Bell Systems Technical Journal*, 57:1475–1488, May-June 1978.
- [11] S. Wei, P. K. Nag, R. D. Blanton, A. Gattiker, and W. Maly. To dft or not to dft? *Proc. International Test Conference*, pages 557–566, 1997.
- [12] T. W. Williams and N. C. Brown. Defect level as a function of fault coverage. *IEEE Transactions on Computers*, C-30:987–988, December 1981.