Evaluation of Effectiveness of Median of Absolute Deviations Outlier Rejection-based I_{DDQ} Testing for Burn-in Reduction

Sagar S. Sabade Duncan M. Walker
Department of Computer Science
Texas A&M University
College Station TX 77843-3112
Tel: (979) 862-4387
Fax: (979) 847-8578
E-mail: {sagars, walker}@cs.tamu.edu

Abstract

CMOS chips having high leakage are observed to have high burn-in fallout rate. I_{DDQ} testing has been considered as an alternative to burn-in. However, increased sub-threshold leakage current in deep sub-micron technologies limits the use of I_{DDQ} testing in its present form. In this work, a statistical outlier rejection technique known as the median of absolute deviations (MAD) is evaluated as a means to screen early failures using I_{DDQ} data. MAD is compared with delta I_{DDQ} and current signature methods. The results of the analysis of the SEMATECH data are presented.

1. Introduction

I_{DDQ} testing can detect many defects that are not detected by stuck-at tests [1]. Its advantages due to current observability surpass advantages of other testing methods and often outweigh its limitations, like slow testing speed [2]. However, leakage current increases as transistor geometries shrink due to reduction in the threshold voltage [3]. Thus rejecting chips that pass other tests but have high I_{DDQ} causes unjustifiable yield loss [4]. Therefore, various techniques have been proposed to extend I_{DDQ} testing to deep sub-micron (DSM) technologies [5, 6, 7, 8, 9]. I_{DDQ} testing is also reported to catch reliability-risk devices and has been shown to reduce the field rejection rate [10]. While complete burn-in elimination may not be possible, burn-in reduction can provide a significant reduction in manufacturing costs.

In this paper we evaluate the effectiveness of I_{DDQ} testing combined with an outlier rejection method for detecting early failures. We use SEMATECH[1] test data for this analysis. The SEMATECH experiment used a chip manufactured with 0.8 μm technology (0.45 μm L_{eff}) as a test vehicle [20]. Therefore, the results presented here will not be valid in the same dimensions for advanced technologies. However, such analysis can be useful to evaluate the effectiveness of these methods by extrapolation of parameters.

This paper is organized as follows. In the next section we describe the motivation for our analysis. In section 3 we outline our methodology. Section 4 describes the results and section 5 concludes the paper.

2. Motivation

IC manufacturers need to ensure that every chip shipped to customers conforms to the specifications. This is normally achieved by testing various parameters against the specifications. Typically, different test patterns are applied at the inputs and outputs are monitored in what is called *functional testing*. Typically CMOS chips have low quiescent current owing to their structure. Thus another way of testing chips (*defect-based testing*) is to measure the leakage current and ensure that it is within acceptable limits. Some chips have defects that are not tested during the entire testing cycle of the chip and therefore fail when they are used in a system. A *field failure* is often costly for customers, especially in applications where high reliability is needed. Early field failures are often the result of manufacturing defects that worsen (to an extent to cause functional failure) due to temperature and/or voltage stress. These failures are called *infant mortality* and form the first part of the *bathtub curve* [2] shown in Fig. 1. Quite often a manufacturer is required to provide a replacement for a customer return, which is an overhead cost for the manufacturer. It might be necessary for a manufacturer to trace back the root cause of the defect and determine whether other parts from the same lot would perform to their specifications.

To reduce customer returns and to maximize profit, manufacturers attempt to catch these early failures before chips are shipped. This is achieved by subjecting chips to high voltage and/or temperature stress in a process called *burn-in*. A burn-in cycle is equivalent to many years of chip

[1] This data comes from the work of the Test thrust at SEMATECH, Project S-121 on Test Methods Evaluation. The results and analysis presented here are our own and do not necessarily represent the views of SEMATECH or its member companies.

operation under normal condition. The stress exerted on a chip causes a defect to accelerate so that it can be detected by a *post burn-in* test. Thus burn-in compresses the time scale of the bathtub curve and reduces time-to-market as shown in Fig. 1. Although effective in screening low reliability chips, burn-in is very expensive. A major cost is the special engineering equipment needed. The time needed for burn-in is another contributing cost factor. Moreover, burn-in is a destructive test [11] and components failing burn-in represent lost revenues. Therefore, manufacturers would like to eliminate this step without increasing the customer reject rate.

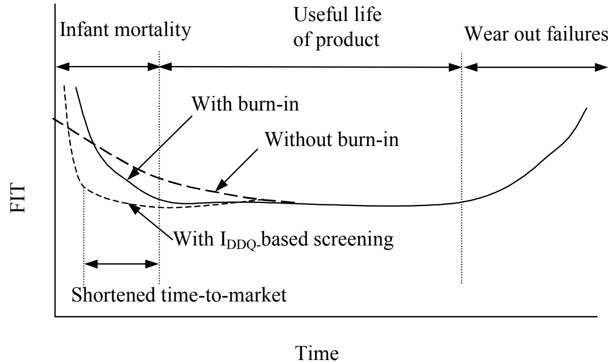

Figure 1. Bathtub curve [2].

It has been shown that I_{DDQ} testing is useful for detecting many defects that lead to reliability hazards. These include gate-oxide shorts, punch through and leaky transistors [12]. In general, chips having higher leakage have higher burn-in fallout rate [13]. I_{DDQ} testing has been used as an alternative to burn-in and some success stories are reported in the literature [14, 15, 16]. Moreover, I_{DDQ} testing takes fraction of the time it would take for burn-in. So considerably shortened time-to-market can be achieved.

As devices geometries shrink it is necessary to reduce the supply voltage (called *constant field* scaling) to maintain a constant electric field [17]. The corresponding reduction in threshold voltage causes exponential increase in the sub-threshold leakage current [5]. Thus it is difficult to distinguish between high defective leakage current and high leakage due to reduced threshold voltage. The traditional method of I_{DDQ} testing where a single current threshold sufficed to differentiate between faulty and fault-free chips is no longer effective. The process variations worsen this fact by causing large chip-to-chip variation in leakage current [18]. Various approaches have been suggested to sustain I_{DDQ} test for DSM technologies [6, 7, 8, 9, 19]. The basic idea is to estimate fault-free I_{DDQ} by considering the effects of process variations, vicinity to faulty chips, radial distance from the center of the wafer, maximum operating frequency, etc.

Statistical techniques are routinely used to find outliers. The chips that have high leakage and are likely to be defective are essentially outliers in the data. In the present study we evaluate the effectiveness of an outlier rejection method to screen early failures. Our goal is to screen outliers in I_{DDQ} data and finds its effect on burn-in fallout rate.

3. Methodology

We use SEMATECH test data for our analysis. The SEMATECH experiment was primarily conducted to evaluate the relative effectiveness of tests [20]. Four types of tests – functional, delay, scan and I_{DDQ} – were performed on 18 466 chips at wafer and package level. SEMATECH experiment used a static threshold of 5 μA for I_{DDQ} test. A sample of devices was subjected to 6, 72 and 144 hours of burn-in and the same tests were performed. It was observed that each test caught unique defects and to a certain extent tests were complementary [21]. The distribution of the test results before burn-in and after six hours of burn-in is shown in Fig. 2. Of particular interest here are the chips that pass all tests and those that fail only I_{DDQ} test. Out of 1558 chips that failed only I_{DDQ} test at wafer level, 1219 failed *only* I_{DDQ} test after burn-in. Fig. 3 shows I_{DDQ} for a sample of chips that passed all the tests at wafer probe. Fig. 4 shows I_{DDQ} for a sample of chips that failed only I_{DDQ} test at wafer probe. Note that since wafer probe and post burn-in tests were conducted at 50°C and 25°C respectively, I_{DDQ} readings are *not* expected to be same for a fault-free chip. For I_{DDQ}-only fail devices the spread in I_{DDQ} values is clearly noticeable in Fig. 4. While some chips do exhibit appreciably increased I_{DDQ} after burn-in (well above the trend line) and are high-risk devices, it would be incorrect to assume that even chips that did not have significant (less than by an order of) increase in I_{DDQ} are defective as well. This is due to the way I_{DDQ} failure was defined in this experiment (single 5 μA threshold). Around 291 of 1558 I_{DDQ}-only failed chips at wafer probe heal after burn-in and move to the "all pass" category after burn-in. They may be high reliability risk devices.

Fig. 5 shows the change in the defect level (DL) and yield loss (YL)/overkill for the corresponding change in the I_{DDQ} threshold. Here overkill and DL were obtained as follows. If a die that passed I_{DDQ} test failed any test except I_{DDQ} after burn-in it is considered to be a failure. Notice that even for a small change in I_{DDQ} threshold, there is considerable change in YL.

Median of Absolute Deviations Outlier Rejection

Many methods for outlier rejection like Chauvenet's criterion [22], Tukey test [23] or Z-scores [24] rely on distribution properties like mean and variance. The presence of outliers in the data causes a shift in the mean and variance. Thus many "true" outliers are not detected. Furthermore, many of these methods assume the data has Normal distribution. A typical I_{DDQ} distribution has a long tail due to outliers. The I_{DDQ} distribution for fault-free chips can be approximated by a lognormal distribution (see Fig. 6). Outliers do not follow any standard distribution.

	IDDQ (5 μA)				
Stuck-at	pass	pass	fail	fail	Delay
pass	1102	79	1558	67	pass
pass	292	43	73	15	fail
fail	16	71	43	141	pass
fail	50	73	217	49	fail
Functional	pass	fail	pass	fail	

(a) Before Burn-in

	IDDQ (5 μA)				
Stuck-at	pass	pass	fail	fail	Delay
pass	1864	6	1367	5	pass
pass	22	0	19	1	fail
fail	9	43	27	85	pass
fail	49	36	254	102	fail
Functional	pass	fail	pass	fail	

(b) After Burn-in

Figure 2. Distribution of chip failures before and after six hours burn-in (Total 3889 chips on 75 wafers).

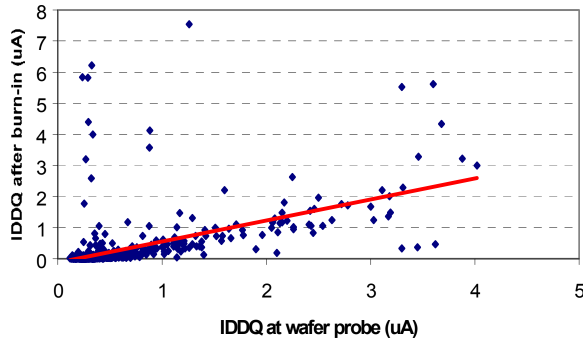

Figure 3. I_{DDQ} before and after burn-in for a sample of chips that passed all tests at wafer probe; chips for which post burn-in $I_{DDQ} < 10$ μA are shown.

For successful outlier detection, we need a resistant estimator that should not be unduly affected by outliers in the sample. The Median of the Absolute Deviations about the median (MAD) is such an estimator [24]. It is defined as:

$$MAD = median_i\left\{\left|x_i - \tilde{x}\right|\right\},$$

where $\tilde{x}$ is the sample median. Then the MAD score (M_i) is defined as:

$$M_i = \frac{0.6745\left(x_i - \tilde{x}\right)}{MAD}$$

The constant 0.6745 is used because for large N for a Normal distribution E(MAD) = 0.6745σ. This M_i is similar to Z-scores. Any observation is labeled as an outlier and rejected when $|M_i| > D$ where D is the maximum permissible MAD score. For large N and Normal distribution a value of 3.5 for D is suggested in the literature [24]. To clarify MAD approach an example of MAD-based outlier rejection is illustrated in the Appendix.

Since outliers do not change the median appreciably, MAD-based rejection has a higher *breakdown point*. The breakdown point of an estimator is defined as the largest proportion of the data that can be replaced by arbitrary values without causing the estimated value to become infinite [24]. The sample mean and standard deviation have breakdown points of zero, as one observation moved to infinity would make these estimators infinite. The sample median has a breakdown point of approximately 50%. The exact percentage depends on whether the number of data points is odd or even.

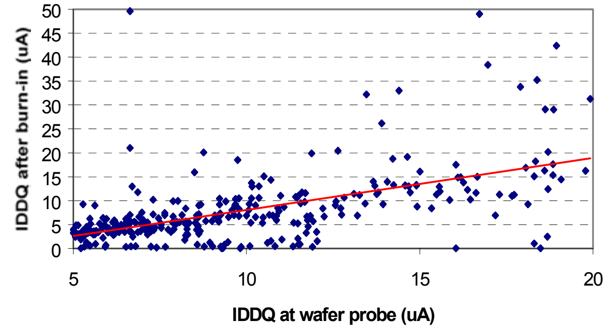

Figure 4. I_{DDQ} before and after burn-in for a sample of chips that failed only I_{DDQ} test at wafer probe; chips for which post burn-in I_{DDQ} is less than 50 μA are shown.

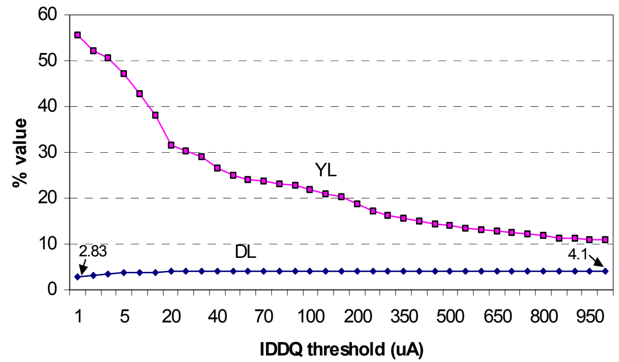

Figure 5. Defect Level and Yield Loss for different I_{DDQ} thresholds, DL is expressed as % of accepted chips and YL is expressed as % of the total chips.

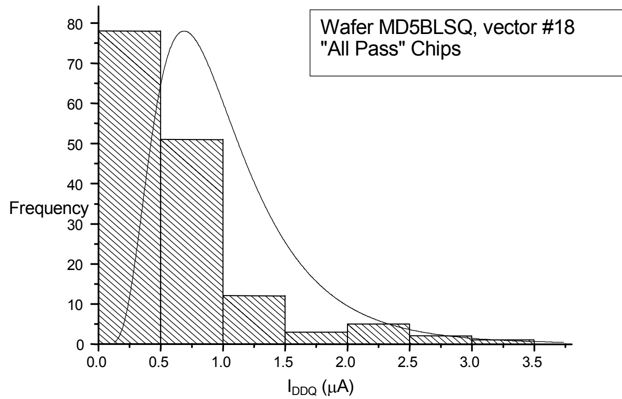

Figure 6. Fault-free I_{DDQ} distribution for a wafer.

MAD Score for I_{DDQ} Testing

Since we consider only chips having six hours burn-in (BI) data, the sample size is reduced to 3889 chips. In practice no I_{DDQ} data is usually available for functional fails. Therefore, wafer level functional and stuck-at fails are screened. The resulting data set contains 3025 chips that have either passed all the tests or failed only I_{DDQ} test at the wafer level. The gross outliers (several mA of I_{DDQ}) are then removed using Chauvenet's criterion [19] with a loose probability threshold of 0.1, following normalizing transform. A different outlier rejection method other than MAD-based rejection is used to avoid bias. This reduces the data set to 2534 chips. For these chips the maximum I_{DDQ} for each vector was less than 500 µA. For each vector a 3σ limit was determined for post BI I_{DDQ} pass/fail decision as described later.

For each die a total of 195 I_{DDQ} measurements are available at probe and after BI. The current industry practice is to use 10-20 I_{DDQ} measurements. We therefore consider only the first 20 measurements for each die. The effective change in overkill and defect level with number of vectors is very small as shown in Fig. 7. This could be because the first few I_{DDQ} vectors are sufficient to detect most of the defects. The I_{DDQ} distribution is converted to a Normal distribution by logarithmic transform as follows. For each vector the minimum nonzero value is found across all chips and all readings are divided by this value. Then logarithm of the ratio is taken. The median I_{DDQ} for each vector is obtained. Then MAD values for each vector are computed. Using these MAD values, MAD scores are computed for each reading.

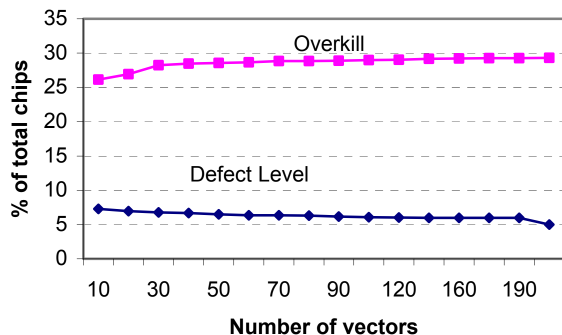

Figure 7: Change in DL and overkill for MAD-based rejection with number of vectors. Voltage fails are ignored from the data set.

Post BI Pass/Fail Decision

If a chip fails any voltage test after burn-in it is considered to be defective. A chip passing all SEMATECH tests after BI, it is considered to be fault-free. Several chips fail only SEMATECH I_{DDQ} test after BI. Considering them all to be defective or defect-free would give misleading results for test escapes and/or yield loss. But how do we decide optimum post burn-in I_{DDQ} threshold? Our solution is to use pre-burn-in I_{DDQ} variation.

We used 3-σ limit obtained from I_{DDQ} at the wafer probe as I_{DDQ} pass/fail threshold for post burn-in data. However, note that because wafer probe was conducted at a higher temperature (50°C) than package level test (room temperature), this limit is rather less stringent. If any I_{DDQ} reading (for the 20 vectors) is more than this limit, a die is considered to be defective. In addition, voltage test failures are also counted as failed chips. There are four possible cases: (a) accepted chip passes all tests after BI, (b) accepted chip fails any test after BI, (c) rejected chip fails any test after BI and (d) rejected chip passes all tests after BI. The case (a) and (c) are correct predictions. The case (b) and (d) are incorrect predictions, the former causing defect level in the shipped lot and the later causing overkill or yield loss. These values are expressed as a percentage of the total number of chips (2534). To determine the sensitivity of the MAD approach to the threshold value, various thresholds were used. Fig. 7 shows defect level and overkill for various thresholds.

The defect level and overkill have an obvious inverse relation. By tightening a control parameter (e.g. MAD threshold), it is possible to increase the quality of the shipped product. The price paid for quality is increased overkill. The advantage in using statistical technique like MAD-based rejection is rejection of only "true" outliers that deviate from the median. This causes a smaller change in overkill for a given quality than a static threshold technique. This can be observed by comparing Fig. 5 and Fig.7. A MAD threshold of 10 was used for the analysis.

4. Experimental Results

To compare the effectiveness of this scheme we performed a similar analysis with delta-I_{DDQ}. Delta I_{DDQ} was defined as the difference between two adjacent readings, thus yielding 19 delta values. A large difference between two consecutive I_{DDQ} indicates likely presence of a defect.

By this definition chips having a passive defect are not rejected if all readings are identical. We computed the mean (μ_δ) and standard deviation (σ_δ) of deltas. If absolute value of any delta exceeded the threshold $\mu_\delta + 3.\sigma_\delta$, the chip was rejected. Current signature [25] is obtained by sorting I_{DDQ} readings. For current signature, we followed a similar approach as in the case of delta I_{DDQ} after sorting the readings.

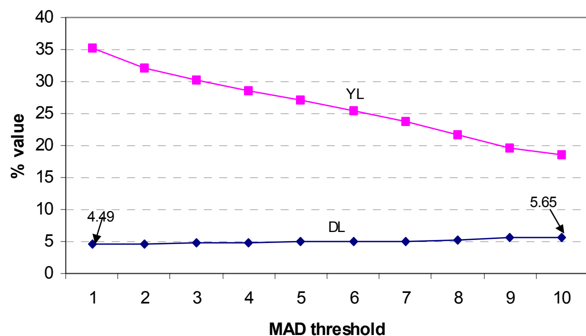

Figure 8: DL and YL for various MAD thresholds, DL is expressed as % of accepted chips and YL as % of total chips. Voltage test fails are ignored from the data set.

Fig. 9 shows the comparison of these methods. The defect level and yield loss values are tabulated in Table 1. DL is expressed as a percentage of accepted chips and yield loss is expressed as a percentage of the total chips.

Table 1: DL and YL for different methods

Method	DL %	YL %
5 μA Threshold	4.2	33.12
Delta I_{DDQ}	19.58	8.88
Current Signature	18.7	60.87
MAD	7.4	14.19

Although the single 5 μA threshold approach has a lower DL, the corresponding yield loss would be unacceptable. Delta-I_{DDQ}, on the other hand, has the highest overall yield but high DL as well. For chips having V_{DD} to ground short, small vector-to-vector variations in I_{DDQ} result in smaller non-zero deltas. Unless a upper static threshold limit is used, these chips are accepted by delta-I_{DDQ}. The reduction in bridge resistance after burn-in would eventually cause increased post-BI I_{DDQ}. On the other hand some of these chips could exhibit healing behavior as shown in Fig. 4. For high reliability requirements, it would be necessary to screen these devices. The MAD scores for these devices would be higher and can reject many of them.

Very low yield of current signature can be explained as follows. After readings are sorted, the mean and variance of deltas reduce. This results in lower pass/fail threshold and rejects many chips that do not exhibit increased leakage after BI thus showing a huge yield loss.

Static-threshold approach is not recommended for BI reduction. The yield loss and defect level figures can vary considerably. MAD-based approach has comparable yield to delta I_{DDQ} with lower defect level. Chips having passive defects have high higher MAD scores for all the vectors and get rejected. MAD is also sensitive to active defects. Since MAD technique is looking at the entire distribution instead of multiple measurements from a single chip, it has higher resolution in screening defects.

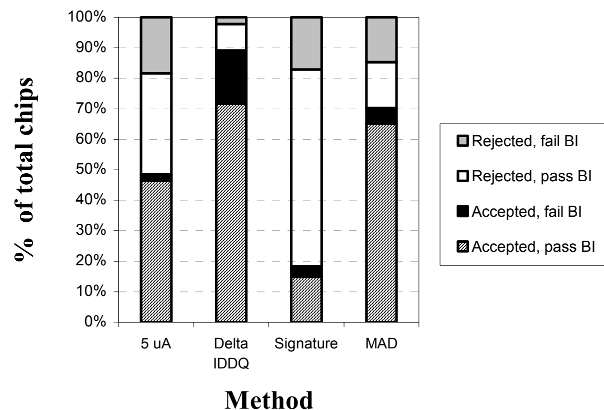

Figure 9: Comparison of various methods

5. Conclusions

The future of single threshold-based I_{DDQ} testing procedure is questionable. The yield loss for this method is unacceptable when achieving DL targets. To account for inter-die and intra-die variations it is necessary to device a methodology that uses different pass/fail threshold for each vector. Statistical outlier rejection techniques can be employed to achieve lower DL. Such techniques can be applied only after the entire data is available. This limits their use in production due to time to market constraints.

When these methods are employed as an alternative to burn-in it would be necessary to use resistant estimators. The outlier rejection methods that employ the mean or the standard deviation are inherently "biased" towards the population. MAD-based and similar techniques that use median or more resistant estimators can eliminate this bias. However, these techniques are not suitable for rejecting chips from a maverick lot. Experiments with more data are needed for fair comparison with other I_{DDQ} test methods.

Acknowledgements

This research was funded in part by the National Science Foundation under grant CCR-9971102. Thanks to Phil Nigh of IBM for providing the SEMATECH data. We would also like to thank anonymous reviewers for their comments and suggestions.

Appendix: MAD Score Computation

This appendix provides an example of MAD score computation and outlier rejection. Consider the data shown in Table 2. Since there are a total of 10 values, we compute the median by averaging the 5th and 6th readings in the ordered data. Thus $\tilde{x}$ = (1+1.01)/2 = 1.005. The fourth column lists the ordered $|x_i - \tilde{x}|$. Thus MAD = (0.025+0.045)/2=0.035. The M_i scores are computed as $0.6745(x_i\text{-}1.005)/0.035$.

Table 1. MAD-based rejection example

No.	Data (x_i)	Ordered Data (x_i)	Ordered $\|x_i - \tilde{x}\|$	M_i
1	1.03	0.76	0.005	0.48
2	0.96	0.89	0.005	-0.87
3	1.11	0.96	0.015	2.02
4	0.76	0.98	0.025	-4.72
5	1.02	1.00	0.025	0.29
6	0.98	1.01	0.045	-0.48
7	0.89	1.02	0.105	-2.21
8	2.34	1.03	0.115	25.72
9	1.01	1.11	0.245	0.09
10	1.00	2.34	1.335	-0.09

The data points 0.76 and 2.34 have M_i values of –4.72 and 25.72 respectively, and so are rejected when using a threshold of 3.5.

References

[1] S. D. McEuen, "I_{DDQ} Benefits," *IEEE VLSI Test Symposium,* Atlantic City, NJ, 1991, pp. 285-290.

[2] R. Rajsuman, "I_{DDQ} Testing for CMOS VLSI," *Proceeding of the IEEE*, Vol. 88, No. 4, April 2000, pp.544-566.

[3] M. Sachdev, "Current-Based Testing for Deep-Submicron VLSIs," *Design and Test of Computers*, March-April 2001, pp. 76-84.

[4] J. Figueras and A. Ferre, "Possibilities and Limitations of I_{DDQ} Testing in Submicron CMOS," *IEEE Transactions on Components, Packaging, and Manufacturing Technology*, part B, Vol. 21, No. 4, Nov. 1998, pp. 352-359.

[5] M. Sachdev, "Deep Sub-micron I_{DDQ} Testing: Issues and Solutions," *European Design and Test Conference,* Paris, March 1997, pp.271-278.

[6] C. Thibeault, "A Novel Probabilistic Approach for IC Diagnosis Based on Differential Quiescent Current Signatures," *VLSI Test Symposium*, Monterey CA, 1997, pp. 80-85.

[7] A. Miller, "I_{DDQ} Testing in Deep Sub-micron Integrated Circuits," *Intl. Test Conference,* Atlantic City, NJ, 1999, pp. 724-729.

[8] P. Maxwell et al., "Current Ratios: A Self-scaling Technique for Production I_{DDQ} Testing," *Intl. Test Conference,* Atlantic City, NJ, 1999, pp. 738-746.

[9] S. Jandhyala et al., "Clustering Based Evaluation of I_{DDQ} Measurements: Applications in Testing and Classification of ICs," *VLSI Test Symposium*, Montreal, Canada, 2000, pp. 444-449.

[10] T. Barrette et al., "Evaluation of Early Failure Screening Methods," *IEEE Intl. Workshop on I_{DDQ} Testing,* Washington D.C., October 1996, pp. 14-17.

[11] R. Rajsuman, *Digital Hardware Testing.* Norwood, MA: Artech House, 1992, ch. 12,pp. 263-295.

[12] J. Soden and C. Hawkins, "I_{DDQ} Testing and Defect Classes – A Tutorial," *IEEE Custom Integrated Circuits Conference*, Santa Clara, CA, May 1995, pp. 633-642.

[13] SEMATECH Report, "Test Method Evaluation: Key Findings and Conclusions," Jan. 1997.

[14] C. Hawkins et al., "Reliability, Test and I_{DDQ} Measurements," *IEEE Intl. Workshop on I_{DDQ} Testing,* Washington D.C., 1997, pp.96-102.

[15] T. Henry and T. Soo, "Burn-in Elimination of a High Volume Microprocessor using I_{DDQ}," *Intl. Test Conference,* Washington D.C., October 1996, pp. 242-249.

[16] S. R. Mallarapu and A. J. Hoffmann, "I_{DDQ} testing on a custom automotive IC," *IEEE J. Solid State Circuits*, Mar. 1995, pp. 295-299.

[17] T. Williams et al., "I_{DDQ} Test: Sensitivity Analysis of Scaling," *Intl. Test Conference,* Washington D.C., October 1996, pp. 786-792.

[18] C. Hawkins and J. Soden, "Deep Submicron CMOS Current IC Testing: Is There a Future?," *Design and Test of Computers*, Oct.-Dec. 1999, pp. 14-15.

[19] S. Sabade and D. Walker, " Improved Wafer-level Spatial Analysis for I_{DDQ} Limit Setting," *Intl. Test Conference,* Baltimore, October 2001, pp. 84-93.

[20] P. Nigh et al. "An experimental study comparing the relative effectiveness of functional, scan, I_{DDQ} and delay-fault testing," *IEEE VLSI Test Symposium,* Monterey CA, 1997, pp. 459-464.

[21] P. Nigh et al., "So what is an optimal test mix? A discussion of the SEMATECH methods experiment," *Intl. Test Conference*, Washington D.C., October 1997, pp. 1037-1038.

[22] J. R. Taylor, *An Introduction to Error Analysis*, University Science Books, 1982.

[23] R. Richmond, "Successful Implementation of Structured Testing," *Intl. Test Conference,* Atlantic City, October 2000, pp. 344-348.

[24] B. Iglewicz and D. Hoaglin, *How to detect and handle outliers,* The ASQC Basic References in Quality Control: Statistical Techniques, Vol. 16, ASQC, 1993, pp.10-13.

[25] A. Gattiker and W. Maly, "Current Signatures: Application," *Intl. Test Conference,* Washington D.C., October 1997, pp. 156-165.