

CircularScan: A Scan Architecture for Test Cost Reduction

Baris Arslan and Alex Orailoglu
Computer Science and Engineering Department
University of California, San Diego
La Jolla, CA 92093
{barslan, alex}@cs.ucsd.edu

ABSTRACT

Scan-based designs are widely used to decrease the complexity of the test generation process; nonetheless, they increase test time and volume. A new scan architecture is proposed to reduce test time and volume while retaining the original scan input count. The proposed architecture allows the use of the captured response as a template for the next pattern with only the necessary bits of the captured response being updated while observing the full captured response. The theoretical and experimental analysis promises a substantial reduction in test cost for large circuits.

1. INTRODUCTION

As the exponential increase in the transistor counts, as predicted by Moore's Law, limits the practicability of sequential and functional patterns, scan-based designs are widely used to reduce test generation cost. Full scan-based design reduces test generation complexity by providing controllability and observability of all the memory elements in the circuit. The test patterns (test responses) are serially shifted in (out) through a scan chain to (from) the memory elements.

Although scan-based designs are able to bound test generation complexity within practicable limits, they come at an increase in test volume and consequently an increase in test application time and a need for higher cost automatic test equipment (ATE) with high pin counts and high memory bandwidth. Current costs of such ATEs range at the levels of thousand dollars per pin [1].

In a traditional scan architecture, the number of scan I/O pins determines the number of scan chains in the circuit as can be seen in Figure 1. As the complexity of the circuit increases, the number of scan cells per scan chain increases, increasing the test application time. On the other hand, an increase in the number of I/O pins necessitates a higher cost ATE with high pin count. A number of approaches have been proposed in the literature to reduce test cost [2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. Generally, the proposed schemes utilize a compression method in conjunction with an on-chip decompression hardware. Run-length coding with cyclical scan chains [2], Huffman code [3] and Golomb code [4] have been utilized to compress the test data. These methods exploit the variable frequencies of test pattern blocks to efficiently store the test data by the decompression being performed on chip. [5] partitions the scan cells into the internal scan chains and all scan chains are loaded in parallel by the same test data; serial load is used for the faults that cannot be detected by the parallel load. [6] improves on [5] by providing two different scan cell partitionings, reducing the number of the serial test patterns. [7] assigns a LFSR to each scan chain and the seeds of the LFSRs are shifted in. [8, 9, 10] reduce the number of scan chains visible to the tester through a decompression network that exploits the low specified bit density of the test

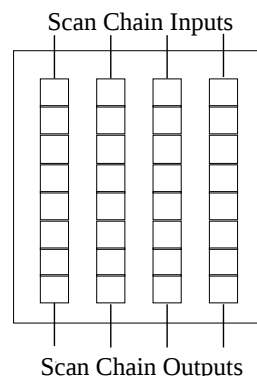


Figure 1: Traditional Multiple Scan Architecture

vectors. [11] encodes the bits that need to be flipped to obtain the next test slice from the current one.

The test generation process generally starts by obtaining a test cube, which has only a few specified bits for each fault. Compaction schemes merge the compatible cubes. The compacted deterministic test patterns generally consist of only a small number of specified bits, 1%-5% as reported in [12]. In practice, the remaining unspecified bits are assigned to random values and fault simulation is performed to drop the accidentally detected faults. In the traditional scan designs, when shifting in the next test pattern, the captured response of the previous pattern is shifted out. Consequently, although only a small percentage of the test pattern bits are specified, this property of the test vector is not exploited and all bits of a test pattern are shifted in.

In this work, a scan architecture is proposed to take advantage of the fact that only a small number of bits is specified in each pattern. The new architecture, while retaining the original number of circuit pins, increases the number of internal scan chains. The captured response of the previously applied pattern is used as a template for the next pattern in the new structure and only the conflicting bits of the next pattern are updated on the captured response of the previous pattern. The proposed approach promises a substantial decrease in test application time and test data for large circuits.

Section 2 discusses the motivation that leads to the method in this paper and introduces the proposed scan architecture for test cost reduction. Section 3 presents the test application process with the proposed architecture. Section 4 provides a theoretical analysis of the test cost reduction with the suggested architecture. Section 5 presents an extensive experimental evaluation and a brief conclusion is offered in section 6.

2. MOTIVATION & SCAN ARCHITECTURE

Deterministic test patterns consist of no more than a small number of specified bits even post-compaction, with the ratio of the specified bits in the 1 to 5 % range in practice [12]. We propose a scan architecture, which allows loading only the specified bits of a pattern quickly instead of loading all bits of a pattern.

In the traditional scan architecture, N pins are employed to feed N scan chains as seen in Figure 1. A compacted test pattern is generated and the unspecified bits are set randomly or by other means, resulting in the test pattern being ready for application to the circuit under test. The completely specified test pattern is loaded to the circuit through the N scan inputs; consequently, the longest scan chain determines the application time of a pattern. Increasing the scan chain count may reduce test application time, but it comes at an increase in the scan input pin count, raising the cost of the ATE.

The proposed architecture exploits the low specified bit density of the test patterns and while limiting the parallel access to all scan chains, it increases the scan chain count in the circuit without changing the pin count. In the preliminary form of the proposed architecture, only one of the scan input pins in the traditional scan architecture is allocated to load the test data. The remaining $N-1$ pins are utilized to select an internal scan chain through a decoder. Since $N-1$ pins can identify 2^{N-1} different scan chains, the number of the scan chains is increased to 2^{N-1} from the original N scan chains in the traditional architecture, yielding a reduction of $2^{N-1}/N$ on the average scan chain length. In the preliminary architecture, all scan chains that have corresponding specified bit(s) in the next pattern are selected in order and loaded with new test data. The captured response, on the corresponding scan chains, to the previously applied pattern is shifted out to a MISR while shifting the new test data in. If the specified bit density is low, since only a limited number of chains has corresponding specified bits, a reduction in the test cost can be delivered by this preliminary form of the proposed architecture. Nonetheless, there are two main problems with this approach. First, since new test data is shifted in to a particular set of scan chains and only their content is shifted out to the MISR, a possible faulty test response on the remaining scan chains may result in undetected faults; a complex and computationally costly analysis is required to preserve the fault coverage and to predict the next vectors under the possibly faulty data. Second, even a single specified bit on a scan chain necessitates loading new test data to all scan cells on the particular scan chain; subsequently, it prevents the utilization of the other scan chains when a scan chain is being loaded even if the next bit to be loaded of the current chain is unspecified.

To overcome the aforementioned shortcomings of the preliminary architecture, it is re-shaped with the following features, re-

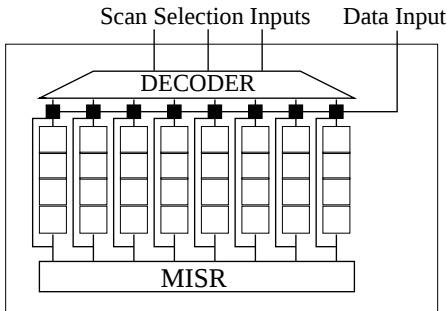


Figure 2: CircularScan: The Proposed Scan Architecture

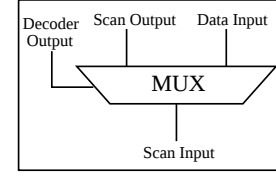


Figure 3: Scan Input Selection Unit

vealing the proposed architecture, denoted as *CircularScan*. First, the output of each scan chain is connected to the input of the same scan chain. Second, although one of the scan chains is selected through the decoder and sourced with new test data as in the preliminary architecture, all other scan chains are also shifted and the scan inputs of these scan chains are sourced with the scan outputs of the same scan chains. Finally, all scan chain outputs are connected to a MISR. The proposed scan architecture, corresponding to the circuit with traditional scan design in Figure 1, can be seen in Figure 2. The *scan selection inputs* identify a scan chain, test data is loaded through the *data input* and scan chain outputs are connected to scan chain inputs through a selection unit. The *scan input selection unit* is depicted in Figure 3. If a particular scan chain is identified by the scan selection inputs through the decoder, the unit feeds the scan chain input with the new data; otherwise, the scan input is sourced by the scan output. Since all scan chains are also connected to the MISR and shift their content, the captured response of the previous pattern is fully observed.

In the proposed architecture, the first test pattern is loaded by selecting each scan chain in order and shifting in the test data through the *data input*, wherein un-specified bits are randomly set. The remaining patterns use the captured response on the scan cells to the previously applied pattern as a template. Only the scan cells with a corresponding specified bit are loaded with new test data if their current contents are at variance with the corresponding specified bits of the pattern.

3. CIRCULARSCAN TEST APPLICATION

In the proposed architecture, the first test pattern is loaded serially. An internal scan chain is selected one at a time through the *scan selection inputs*. The test data for the selected chain, wherein the unspecified bits are randomly set, is shifted in through the *data input*.

The remaining patterns use the captured response on the scan cells to the previously applied test pattern as a template. The content of a scan cell is changed if it corresponds to a specified bit of the next pattern to be applied and if its current content is at variance with the particular specified bit.

The bits of a test pattern, corresponding to the scan cells that have equal distance to the scan inputs, constitute a test slice. In the current test slice, if a set of bits differs from the corresponding bits in the scan cells, one of them is selected through the decoder and the value in the scan cell is updated through the *data input*. Since only one of the conflicting bits of the current test slice can be changed at each cycle, multiple full rotations of the scan chain content may be required. Therefore, the test slice with the highest number of conflicting bits of a test pattern determines how many times the scan chains have to be fully rotated to apply the current pattern. In the first rotation, the captured response of the previous pattern is fed to the MISR, thus delivering full observation of the test responses.

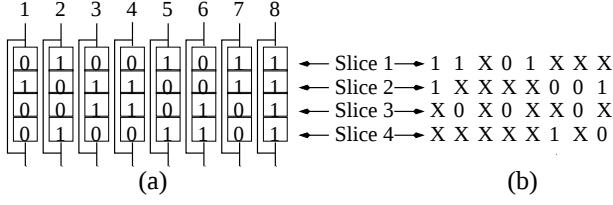


Figure 4: Captured Test Response and Next Pattern

Example Assume that the captured response for the previous test pattern is given in Figure 4 (a) and the next test pattern to be applied is given in Figure 4 (b). If the traditional scan architecture in Figure 1 is employed, the application of the next test pattern is performed by shifting the test pattern through the 4 input pins in a total of 8 cycles. If *CircularScan* is used, there are only 4 specified bits of the current pattern that vary from the captured response of the previously applied pattern. Consequently, only updating these particular bits with the new architecture is sufficient to load the next pattern. In the first cycle, chain 8 is selected and logic 0 is shifted to the selected chain through the data input. Other chains also shift one cycle and the values on the chain outputs are fed to the chain inputs. The content of the scan cells after the first cycle is shown in Figure 5 (a). Chain 4, chain 7 and chain 1 are selected in the given order for the consequent three cycles and logic 0, 0 and 1 are shifted in through the data input, respectively. The content of the scan cells is shown in Figure 5(b)-(d) after the cycles 2-4, respectively. Consequently, in this example, only 4 cycles are necessary to apply the next pattern and the unspecified bits are pseudo-randomly filled with the response of the previous pattern.

4. TEST APPLICATION COST ANALYSIS FOR CIRCULARSCAN

Test application time and test data volume depends on the number of test patterns, T , the number of input pins, I , the number of scan cells, N , and the scan frequency, F . In the traditional scan architecture, if perfectly balanced scan chains and a scan chain per input pin is assumed, the test application time, t , and test data volume, v , can be found by the following equations:

$$t = T \times (\lceil N/I \rceil + 1) \times F^{-1} \quad (1)$$

$$v = T \times \lceil N/I \rceil \times I \quad (2)$$

In the proposed architecture, the number of scan chains, S , that can be supported is increased to 2^{I-1} even as the number of the pins necessary is kept constant. Since the first test pattern is loaded serially, N cycles are required to apply it. The application time of the remaining patterns is determined by the test slice with the highest number of conflicting bits. Formally, let k_j be the number of bits that conflict with the captured test response in the j th test slice and let K_{max} be the maximum of k_j 's. The scan content is fully rotated K_{max} times to apply the next pattern and each rotation requires $\lceil N/S \rceil$ cycles.

If the specified bit density of the test set is p , a single bit of a test vector exhibits a conflicting value to the captured response bit with a probability of $p/2$. The probability of having i conflicting bits in the j th test slice ($P(k_j = i)$) is given by the following equation:

$$P(k_j = i) = \binom{S}{i} \left(\frac{p}{2}\right)^i \left(1 - \frac{p}{2}\right)^{S-i} \quad (3)$$

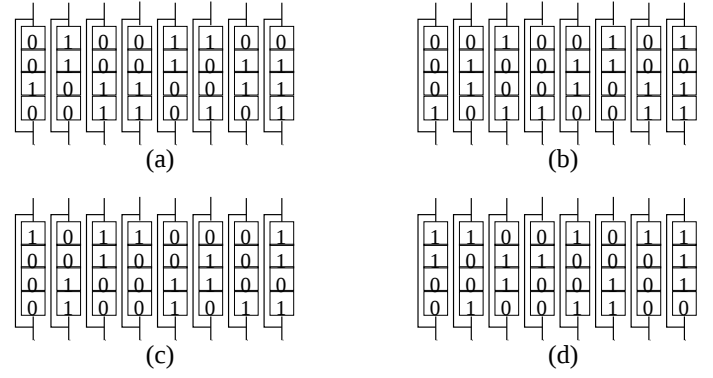


Figure 5: Application of the Next Pattern

Given a total of N scan cells and S scan chains, the total number of test slices is $\lceil \frac{N}{S} \rceil$. Subsequently, the probability of K_{max} being equal to i (i.e., $P(K_{max} = i)$) can be found by the following equations:

$$P(K_{max} = i) = (P(k_j \leq i))^{\lceil \frac{N}{S} \rceil} - (P(k_j \leq i-1))^{\lceil \frac{N}{S} \rceil} \quad (4)$$

$$P(k_j \leq m) = \sum_{k=0}^m \binom{S}{k} \left(\frac{p}{2}\right)^k \left(1 - \frac{p}{2}\right)^{S-k} \quad (5)$$

Assuming the specified bits are uniformly distributed, the number of rotations needed to apply a test pattern is, on average, the expected value of the K_{max} ($E(K_{max})$):

$$E(K_{max}) = \sum_{i=1}^S i \times P(K_{max} = i) =$$

$$S - \sum_{i=0}^{S-1} \left(\left(\sum_{k=0}^i \binom{S}{k} \left(\frac{p}{2}\right)^k \left(1 - \frac{p}{2}\right)^{S-k} \right)^{\lceil \frac{N}{S} \rceil} \right) \quad (6)$$

Consequently, for the specified bit density of p , the test application time, t_s , and test data volume, v_s , with *CircularScan* can be found by the following equations:

$$t_s = (N + (T-1) \times (E(K_{max}) \times \lceil N/S \rceil + 1)) \times F^{-1} \quad (7)$$

$$v_s = N + (T-1) \times E(K_{max}) \times \lceil N/S \rceil \times I \quad (8)$$

Given the equations (1), (2), (7) and (8), the reductions in test application time and test data volume can be obtained by the equations (9) and (10), respectively.

$$t_r = 1 - \left(\frac{t_s}{t}\right) \quad (9)$$

$$v_r = 1 - \left(\frac{v_s}{v}\right) \quad (10)$$

The provided theoretical analysis of the test application time and test data volume assumes a uniform distribution of the specified bits in a test vector. In practice, the scan cells of a particular scan chain are usually driven by the same logic. As a result, the specified bits only appear on a few scan chains instead of as a uniform distribution; it can be subsequently expected that the results of the proposed architecture may exceed even the strong results predicted by the provided mathematical analysis shown here.

	$I=6$		$I=7$		$I=8$		$I=9$	
Circuit	t_r	v_r	t_r	v_r	t_r	v_r	t_r	v_r
s13207	54.6	55.0	63.6	64.1	65.5	66.2	68.0	68.8
s15850	29.8	29.8	37.8	38.0	43.5	43.8	43.0	43.3
s35932	6.2	4.0	17.2	15.1	19.9	17.8	22.7	20.7
s38417	25.4	25.9	35.9	35.8	43.3	43.4	41.8	41.8
s38584	32.1	32.0	43.3	43.3	51.5	51.6	55.1	55.3

Table 1: Test Application Time and Test Volume Reduction

5. EXPERIMENTAL RESULTS

The proposed scan architecture exploits the low specified bit density of the test vectors to reduce test data volume and test application time. The new scan architecture exponentially increases the number of scan chains while retaining the original scan input count. Consequently, the test cost savings of the proposed method can be observed best in circuits with a high gate and scan input pin counts.

5.1 Experimental Framework

The performance of the proposed method has been analyzed with the larger ISCAS89 [13] benchmark circuits. The ATALANTA test generation tool [14] and the HOPE fault simulation tool [15] have been used for the experiments. The general test information regarding the circuits can be seen in Table 2; the second column denotes the number of scan cells, the third column denotes the number of test vectors, the fourth column denotes the number of faults and the last column denotes the specified bit density (p) of the test vectors on average.

The compaction algorithm employed in this work is shown in Figure 6. The algorithm is a slightly modified version of the one presented in [9]. The algorithm selects a seed fault from the fault list and generates a test cube for this particular fault. The fault list is traversed and the test cubes of the faults that are compatible with the seed are merged with the seed cube. Whenever a fault that is not compatible with the seed is encountered, the fault counter is incremented. When the fault counter exceeds a predetermined limit or the fault list is exhausted, the unspecified bits of the current combined cube are filled by the corresponding response bits of the previous test vector¹ and fault simulation is performed. The detected faults are dropped from the fault list and the algorithm continues by selecting a new seed fault from the fault list.

5.2 Experimental Data Evaluation

Table 1 lists the results of the application of the proposed method to the benchmark circuits. In Table 1, the second, third, fourth and fifth columns denote the test application time reduction (t_r) and test

¹If the current cube is the first test vector, the unspecified bits are set randomly.

Circuit	Flop#	T	f	$p(\%)$
s13207	669	299	9664	5.6
s15850	597	186	11336	11.3
s35932	1728	34	35110	10.3
s38417	1636	270	31015	14.1
s38584	1452	251	34797	7.7

Table 2: General Circuit Information

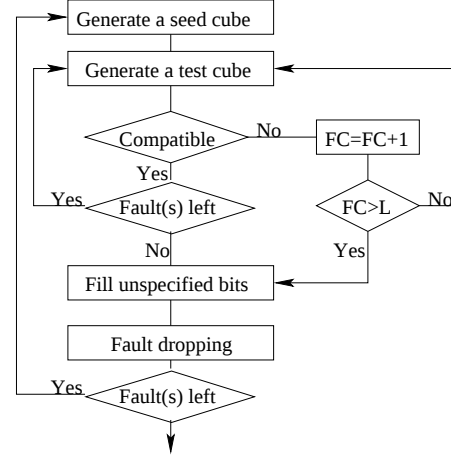
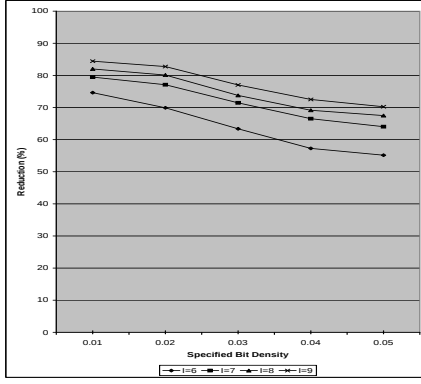


Figure 6: Compaction Algorithm

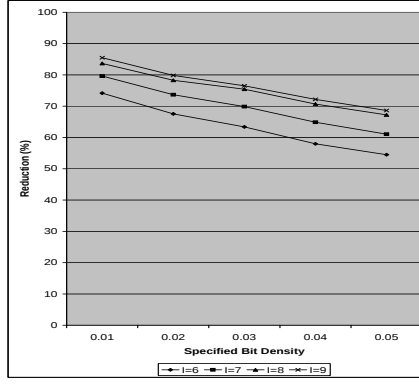
data volume reduction (v_r) for scan input counts of 6, 7, 8 and 9, respectively. Even the larger ISCAS89 benchmark circuits are relatively quite small in comparison to current industrial circuits and their small fault lists allow a compatibility check among a large set of test cubes in a feasible time during the compaction process. As can be seen from the last column of Table 2, the specified bit density of the compacted test vectors is high in comparison to the reported specified bit density of the industrial test vectors. Since the proposed architecture is designed to exploit the low specified bit density of the test set, the test cost reduction obtained for ISCAS89 circuits as presented in Table 1 is not a strong indication of the expected test cost reduction for the much larger industrial circuits typical of the state of the art nowadays. It can be expected that the proposed architecture will provide better test cost reduction for the larger circuits. The increase in the test cost reduction as the specified bit density decreases in Table 1 supports this expectation for the large circuits with low specified bit density of the test vectors.

The most important feature of the proposed architecture is that it increases total scan chain count exponentially; 2^{I-1} scan chains instead of the original I scan chains. Since the benchmark circuits have a small number of scan cells, the results in Table 1 are reported up to the input pin count of 9 only. Since the proposed method exponentially increases the scan chain count, it will provide a much improved reduction in test cost for the larger scan input counts; the steady increase of the test cost reduction in the provided results as the scan input count gets larger is a strong indication of the predicted improvement in test cost.

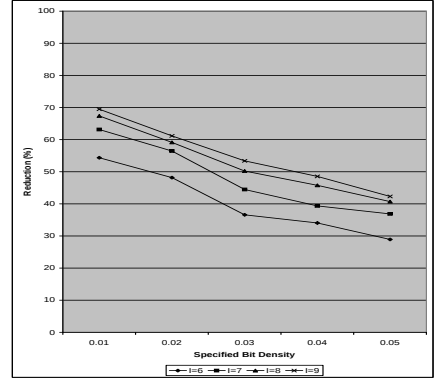
In the second set of experiments, in order to explore the possible test cost reduction by the proposed architecture for lower speci-



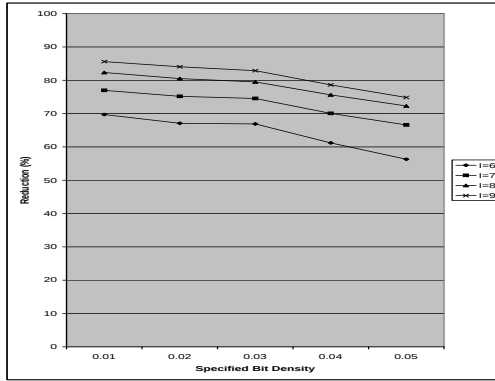
(a) S13207



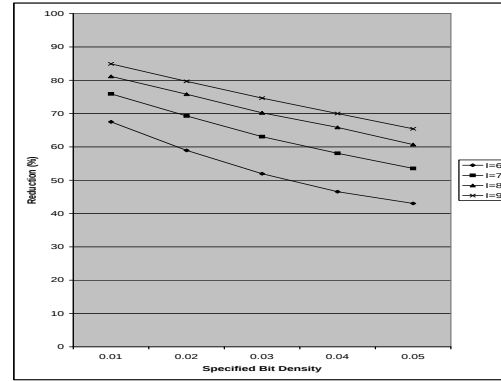
(b) S15850



(c) S35932



(d) S38417



(e) S38584

Figure 7: Test Time Reduction

fied bit densities, the compaction algorithm is slightly modified to stop the compaction process when the specified bit density of a test vector reaches a predetermined value. The graphs in Figure 7 provide the test application time reduction² for the specified bit density range of 1%-5% and for the scan input counts of 6-9. As depicted in the graphs, the scan input count increase and the specified bit density decrease deliver a steady increase in the test application time reduction. The reduction moves up to 85-90% levels for the lower specified bit densities. The reported test cost reduction even for the small input counts of 6-9 is very promising for the possible reduction in the circuits with larger scan input counts.

In order to validate the theoretical analysis that has been presented in Section 4, for the same circuits and test sets that have been used in the experimental results of Figure 7, the theoretically predicted test application time reductions, calculated by equation (9), are provided in Figure 8. As the comparison of the results in Figure 7 and Figure 8 indicates, the actual behavior and the theoretically expected behavior are very similar. The actual reduction is slightly better than the theoretically obtained reduction for most of the cases as a result of the pessimism inherent in the uniform specified bit distribution assumption in the theoretical analysis. Consequently, the provided analysis can be used safely to estimate the expected reduction for a given circuit and test set.

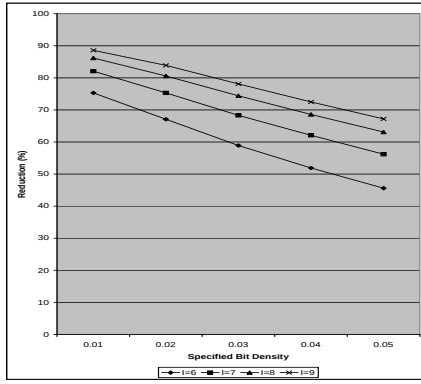
²Since test application time reduction and test volume reduction is similar, only the results for test application time reduction are provided.

6. CONCLUSIONS

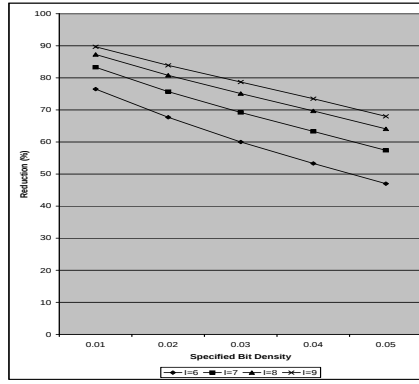
As the circuits get larger, the scan-based designs are used extensively to reduce the complexity of the test generation process. Although scan-based design offers practical test generation times, it comes at an increase in test application time and test data volume. A new scan design, *CircularScan*, has been proposed in order to reduce the test application time and test data volume with no necessity for high cost ATEs.

In practice, even compacted test cubes are composed of a small number of specified bits. The proposed architecture is designed to exploit the low specified bit density of the test vectors. Instead of loading all bits of a test vector, the response to the previously applied vector is used as a template and only the bits that are at variance with the next pattern are updated. *CircularScan* increases exponentially the scan chain count in the circuit while retaining the original scan input pin count. One of the varying bits of a test slice in the new scan chains is updated for each shift cycle; achieved by sourcing one of the scan chains through the data input and by feeding the rest of the scan chains by their own output.

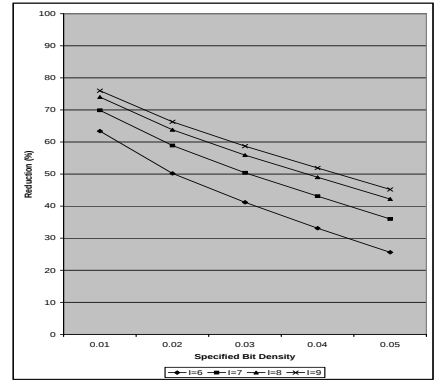
A mathematical analysis and associated experimental results indicate that a substantial reduction can be achieved in test application time and test data volume. The proposed method is not only easily scalable to larger industry circuits but the experimental results and the theoretical analysis further predict even higher test cost reduction for the larger circuits.



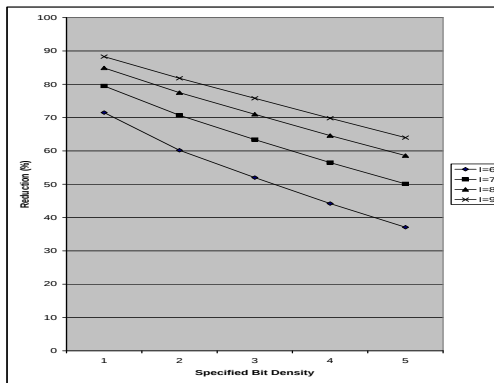
(a) S13207



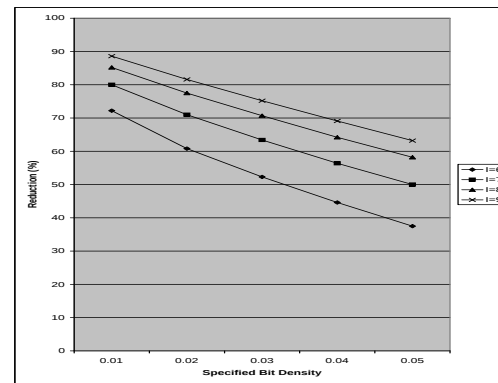
(b) S15850



(c) S35932



(d) S38417



(e) S38584

Figure 8: Theoretical Test Time Reduction Analysis

7. REFERENCES

- [1] V. Agrawal and M. Bushnell, *Essentials of Electronic Testing for Digital, Memory and Mixed-Signal VLSI Circuits*, Kluwer Academic Publishers, 2000.
- [2] A. Jas and N. Touba, "Test Vector Decompression via Cyclical Scan Chains and Its Application to Testing Core-Based Designs", in *ITC*, pp. 458–464, 1998.
- [3] A. Jas, J. Ghosh-Dastidar and N. Touba, "Scan Vector Compression/Decompression Using Statistical Coding", in *VTS*, pp. 25–29, 1999.
- [4] A. Chandra and K. Chakrabarty, "System-on-a-chip Test-Data Compression Architectures Based on Golomb Codes", *IEEE TCAD*, vol. 20, n. 3, pp. 355–368, March 2001.
- [5] I. Hamzaoglu and J. Patel, "Reducing Test Application Time for Full Scan Embedded Cores", in *FTCS*, pp. 260–267, 1999.
- [6] A. R. Pandey and J. H. Patel, "Reconfiguration Technique for Reducing Test Time and Test Data Volume in Illinois Scan Architecture Based Designs", in *VTS*, pp. 9–15, 2002.
- [7] A. Jas, B. Pouya and N. Touba, "Virtual Scan Chains: A Means for Reducing Scan Length in Cores", in *VTS*, pp. 73–78, 2000.
- [8] I. Bayraktaroglu and A. Orailoglu, "Test volume and application time reduction through scan chain concealment", in *DAC*, pp. 151–155, 2001.
- [9] I. Bayraktaroglu and A. Orailoglu, "Decompression Hardware Determination for Test Volume and Time Reduction through Unified Test Pattern Compaction and Compression", in *VTS*, pp. 113–118, 2003.
- [10] W. Rao, I. Bayraktaroglu and A. Orailoglu, "Test Application Time and Volume Compression Through Seed Overlapping", in *DAC*, pp. 732–737, 2003.
- [11] S. Reda and A. Orailoglu, "Reducing Test Application Time Through Test Data Mutation Encoding", in *DATE*, pp. 387–393, 2002.
- [12] T. Hiraide, K. O. Boateng, H. Konishi, K. Itaya, M. Emori and H. Yamanaka, "BIST-Aided Scan Test - A New Method for Test Cost Reduction", in *VTS*, pp. 359–364, 2003.
- [13] F. Brglez, D. Bryan and K. Kozminski, "Combinational Profiles of Sequential Benchmark Circuits", *IEEE Int. Symp. on Circuits and Systems*, vol. 3, pp. 1929–1934, May 1989.
- [14] H. K. Lee and D. S. Ha, *On the Generation of Test Patterns for Combinational Circuits*, Technical Report 12-93, Department of Electrical Eng., Virginia Polytechnic Institute and State University.
- [15] H. K. Lee and D. S. Ha, "HOPE: An Efficient Parallel Fault Simulator", in *DAC*, pp. 336–340, 1992.