

2009

# A Hybrid Test Architecture to Reduce Test Application Time in Full Scan Sequential Circuits

Priyankar Ghosh

*Indian Institute of Technology Kharagpur, priyankar@cse.iitkgp.ernet.in*

Srobona Mitra

*Indian Institute of Technology Kharagpur, srobona@cse.iitkgp.ernet.in*

Indranil Sengupta

*Indian Institute of Technology Kharagpur, isg@cse.iitkgp.ernet.in*

Bhargab B. Bhattacharya

*Indian Statistical Institute, Kolkata, India, bhargab@isical.ac.in*

Sharad C. Seth

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

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



Part of the [Computer Sciences Commons](#)

---

Ghosh, Priyankar; Mitra, Srobona; Sengupta, Indranil; Bhattacharya, Bhargab B.; and Seth, Sharad C., "A Hybrid Test Architecture to Reduce Test Application Time in Full Scan Sequential Circuits" (2009). *CSE Conference and Workshop Papers. 2.*

<http://digitalcommons.unl.edu/cseconfwork/2>

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.

# A Hybrid Test Architecture to Reduce Test Application Time in Full Scan Sequential Circuits

Priyankar Ghosh, Srobona Mitra,  
and Indranil Sengupta  
Indian Institute of Technology Kharagpur  
Kharagpur-721302, India  
{priyankar,srobona,ising}@cse.iitkgp.ernet.in

Bhargab Bhattacharya  
Indian Statistical Institute  
Kolkata-700108, India  
bhargab@isical.ac.in

Sharad Seth  
University of Nebraska-Lincoln  
Lincoln NE 68588-0115, U.S.A.  
seth@cse.unl.edu

**Abstract**—Full scan based design technique is widely used to alleviate the complexity of test generation for sequential circuits. However, this approach leads to substantial increase in test application time, because of serial loading of vectors. Although BIST based approaches offer faster testing, they usually suffer from low fault coverage. In this paper, we propose a hybrid test architecture, which achieves significant reduction in test application time. The test suite consists of: (i) some external deterministic test vectors to be scanned in, and (ii) internally generated responses of the CUT to be re-applied as tests iteratively, in functional (non-scan) mode. The proposed architecture uses only combinational ATPG to hybridize deterministic testing and test per clock BIST, and thus makes good use of both scan based and non-scan testing. We also present a bipartite graph based heuristic to select the deterministic test vectors and sequential fault simulation technique is used to perform the exact analysis on detected faults during the re-application of internally generated responses of the CUT during testing. Experimental results on ISCAS-89 benchmark circuits show the efficacy of the heuristic and reveal a significant reduction of test application time.

**Keywords** - ATPG, DFT, BIST, LFSR, and MISR

## I. INTRODUCTION

One of the disadvantages of using scan based design is the huge test application time due to scan operations. Reducing the test application time is one of the critical factors to reduce the test cost for a scan circuit [1].

Dynamic compaction based approaches [2], [3], [4] strive to obtain highly compact test set. Static compaction techniques [5], [6] try to reduce the total number of scan operations. Limited scan operations [7], [8] only shift in new values for a subset of the flip-flops, and at the same time shift out the values stored in those flip-flops. An approach [9] tries to reduce the useless patterns generated by LFSRs. Another approach [10] focuses to find a minimum set of segments in the LFSR sequence, where each segment corresponds to a consecutive subsequence of useful test patterns. Another design-for-testability (DFT) technique employed to eliminate test data volume and to reduce test application time in a synchronous sequential circuit is the use of *autoscan* [12], [13], which integrates both the functional mode and the DFT mode in the test process.

In this paper, a hybrid test architecture is proposed to reduce the test application time. The method combines deterministic testing and BIST techniques. The test suite uses external

deterministic test vectors that are applied in the scan mode, as well as internally generated responses of the CUT which are re-applied as tests over several iterations, in non-scan mode. The idea of using response of a test vector as a second vector was used earlier in [11] to generate a test pair for detecting a delay fault.

We elaborate on the basic concept using a small example in Section II. Section III describes the proposed test architecture. The heuristic for generating the test suite is presented in Section IV. The experimental results of this simulation are presented in Section V.

## II. MOTIVATING EXAMPLE

In this section the basic procedure is illustrated using a small example. Let us consider the ISCAS-89 benchmark circuit s27 which has 4 inputs, 1 output, 3 flip-flops and 10 gates. ATALANTA generates 6 test patterns for the combinational part of s27. In order to apply each pattern, 3 clock cycles are required to shift in the pattern into the scan chain. We call these vectors *scan test vectors* because these vectors are shifted in or scanned out of the scan chain.

The motivation behind this work is to reduce the test application time by cutting down the number of scan test vectors. Internally generated responses are used as test vectors in order to compensate the fault coverage. On applying a scan test vector, the part of the response that is stored in the flip-flops, is applied at the next clock directly to the pseudo primary inputs of the CUT. An LFSR is used to generate pseudo-random patterns for the primary inputs. Therefore the next test vector is created using the response of the previous vector and the pseudo-random pattern generated by the LFSR which is applied in the next clock. These circulations are done per clock basis, thereby saving a lot of time.

The circulation of the response is continued as long as it detects new faults. When the circulated response fails to detect any new fault among the remaining faults, the circulation is stopped. Experimental results show that even if the circulation of response is continued, new faults are hardly detected.

Table I shows the circulation of response. In this approach we are able to reduce the number of scanned test vectors by 3. So we save 9 clock cycles of test application time. Instead

PI	PPI	Application Procedure	PO	PPO	Activity
0000	000	Scan In	1	000	PPI of next vec.
0111	000	Circulate response	0	000	-do-
1101	000	-do-	1	101	-do-
1000	101	-do-	1	101	Scan Out
1001	000	Scan In	0	010	PPI of next vec.
0000	010	Circulate response	0	010	Scan Out
0100	011	Scan In	0	011	Scan Out
0001	111	Scan In	1	001	Scan Out

TABLE I  
CIRCULATING THE RESPONSE FOR S27

of that we circulated the response 4 times. So we reduce the total test application time by 5 clock cycles.

### III. THE PROPOSED TEST ARCHITECTURE

Since the design is full scan based, flip-flops form a single or multiple scan chain(s) depending on the configuration. The proposed architecture (Figure 1) consists of the following modes three modes:

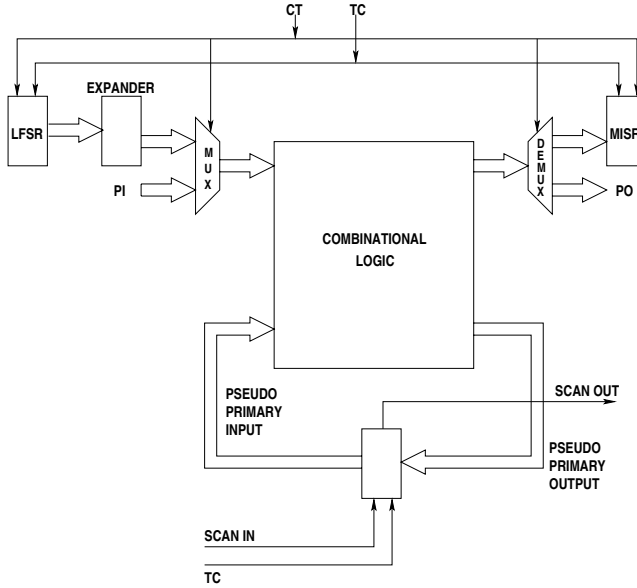


Fig. 1. Proposed test architecture

- 1) *Normal mode* : This mode is used for operating the circuit according to its functional specification.
- 2) *Scan mode* : In this mode the following operations take place:
  - The scan test vectors are shifted into the scan chain.
  - Initial patterns are shifted into the LFSRs.
  - The responses that are stored in scan chain and MISR are shifted out.
- 3) *Circulating response mode* : In this mode, the pseudo random patterns generated by the LFSR are applied to the primary inputs for testing. In this mode only the

LFSR will be activated while the flip-flops will behave as in the normal mode.

Since the number of primary inputs can be of the order of hundreds, an ‘expander’ is used to provide pseudo random patterns to a large number of primary input lines using a small LFSR. Every output line of the expander circuit is driven by an XOR gate. The inputs of the XOR gate are connected to the outputs of the flip-flops of the LFSR in a round robin fashion. The construction of the ‘expander’ is as follows. Suppose the LFSR has  $k$  bits and there are  $n$  primary inputs and  $X_i$  is the ‘XOR’ gate that drives the  $i^{th}$  primary input. The  $j^{th}$  flip-flop of the LFSR is connected to  $X_\varphi^1$  and  $X_\psi^2$  where  $1 \leq \varphi \leq n$ ,  $2 \leq \psi \leq (n + 1)$ ,  $\varphi \bmod m = j$ , and  $(\psi - 1) \bmod m = j$ .  $X_m^1$  and  $X_m^2$  represent the first and the second input of the  $m^{th}$  ‘XOR’ gate respectively. Figure 2 shows the schematic of the expander. A multiplexer is used to select between the input from LFSR and the external input that comes from the input pins.

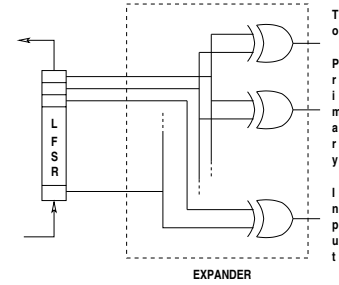


Fig. 2. Schematic of expander

Using two input lines TC and CT, three modes are implemented. During normal mode of operation both TC and CT are kept at 0. In scan mode (TC = 1, CT = 0) the flip-flops form a scan chain, and shift-in, shift-out operations take place. The inputs to the flip-flops come from the combinational part of the circuit when TC is 0. In circulating response mode (TC = 0, CT = 1) the LFSR gets activated and primary inputs are driven using the pseudo random patterns generated by it. The clock drives the LFSR to the next state, thus the next pseudo random pattern for the next circulating vector is generated. The initial state of the LFSR is loaded during the scan mode. Also the pattern stored in the MISR is shifted out in the scan mode.

The CT line also controls the multiplexer. In normal mode of operation the multiplexer allows the primary input to come directly from the input pins, whereas in circulating test mode of operation the primary input is driven by the output of the expander. In the scan mode neither does the primary input have any effect on the state of the flip-flops, nor do the patterns corresponding to the primary input participate in testing. However, during the scan mode CT is kept at 0. The demultiplexer is also controlled similarly.

A Multiple Input Signature Register (MISR) is added to compress the primary output values during the circulation of responses. A scan test vector and its subsequent sequence

of circulated responses are applied and the corresponding primary output patterns are compressed into the MISR. The compressed signature is observed before the application of the next scan test vector.

#### IV. TEST VECTOR GENERATION

To select the scan test vectors we use a two step process. In the first step, a bipartite graph is created, where one independent set represents the faults and the other independent set represents the test vectors. In the second step, the scan test vectors are selected from the graph.

##### A. Generation of Bipartite Graph

A bipartite graph is  $G = \{V, E\}$ , where  $V$  is set of vertices and  $E$  is the set of edges. Let  $TV$  represent the test vectors,  $FL$  represent the set of faults.  $TV$  and  $FL$  are the independent sets of graph  $G$ . Each edge  $e_{ij} \in E$  between vertex  $t_i \in TV$  and  $f_j \in FL$  indicates that vector  $t_i$  detects the fault  $f_j$ .

---

##### Algorithm 1: Bipartite Graph Generation

---

```

1 Initialize the degrees of all  $v_i \in V$  to 0 ;
2 foreach each test vector  $t_i \in TV$  do
3   foreach each fault  $f_i \in FL$  do
4     Perform concurrent fault simulation;
5     if  $t_i$  detects  $f_i$  then
6       Create an edge  $e_{ij}$  between  $t_i$ ;
7     end
8   end
9 end

```

---

##### B. Heuristic to Select Scan Test Vectors

Scan test vectors are selected from  $G$  using a greedy approach. The maximum degree vertex  $t_i \in TV$ , that is, the vector which detects the maximum number of faults is selected as scan test vector. The test vector  $t_i$  is  $\langle pi_i \rangle \langle ppi_i \rangle$ , where  $\langle pi_i \rangle$  is the pattern corresponding to the primary inputs and  $\langle ppi_i \rangle$  is shifted into the scan chain. The response of test vector  $t_i$  is  $r_i = \langle po_i \rangle \langle ppo_i \rangle$ , where  $\langle po_i \rangle$  is the valuation of the primary outputs and  $\langle ppo_i \rangle$  is the pattern stored into the scan chain and typically shifted out of the scan chain. In this work,  $\langle ppo_i \rangle$  is used as  $\langle ppi_{i+1} \rangle$ . The  $\langle pi_{i+1} \rangle$  is set using the LFSR. The next test vector  $\langle pi_{i+1} \rangle \langle ppi_{i+1} \rangle$  is applied in the next clock, thereby the scan shift operations are avoided.

Sequential fault simulation is done to compute which faults are detected by the scan test vector and its subsequent circulations. The details of the sequential fault simulation algorithm is provided later. When the circulations fail to detect any new fault for a certain specified length  $C_{len}$ , the circulation of response is stopped. The reason for continuing the circulation is discussed later. At the end of each circulation, the detected faults and the edges incident on them are deleted from  $G$ .

Again the maximum degree vertex  $t_i \in TV$  is selected as the next scan test vector and the circulation length of  $t_i$  is computed similarly. This process of selecting test vectors for scan operation and generation of its subsequent circulation is continued till the desired fault coverage is achieved.

To perform the sequential fault simulation, we implemented a modified version of the existing concurrent fault simulation algorithm. Typically at the end of the application of the concurrent fault simulation algorithm, each primary output(PO) line and pseudo primary output(PPO) line contains a list of faults(bad-gates) that are detected by the current vector. Since the pattern corresponding to the primary outputs is stored in the MISR, the faults that are detected at primary outputs are dropped from the fault list immediately. Also the vertices that represent these faults along with the edges incident on them are deleted from  $G$ . Since the pattern corresponding to the PPOs are not observed immediately, the fault-lists corresponding to the PPO lines are not dropped at that point of time. At the next cycle, the pattern corresponding to the PPOs is re-applied to the PPIs. The fault-lists are added to the corresponding PPIs and the fault simulation is carried out with the response of the previous test vector and the pattern generated by the LFSR. At the end of each circulation, the pattern stored in the scan chain is shifted out. Therefore at the end of circulation, the list of faults corresponding to the PPOs will be definitely detected. Then those faults are dropped and  $G$  is updated by removing the vertices that represent these faults along with the edges incident on them.

We have observed that even if a particular circulation fails to detect any new fault, the next circulation may detect new faults. Thus the circulation of response is continued for a certain number  $C_{len}$  and if the sequence fails to detect any new fault we select the next scan test vector from  $G$ . It is also found from experimental results that the improvement depends on the sequence length and significant improvement can be achieved for large circuits. We have carried out experiments using different sequence length and the result is given in the next section.

It is observed when the fault coverage approaches the target value, faults that remain in the fault list are relatively harder to detect. In other words when the fault coverage approaches desired level, more and more vectors are encountered whose immediate response does not detect any new fault. This problem is addressed in the following way during the selection of the next scan test vector. For those vectors whose immediate response fails to detect any new fault, a flag is set when they are selected for the first time. First the maximum degree vertex  $t_i \in TV$  is selected and if its response does not detect any new fault, the next best vector among the untried vector set is selected. After all vectors are tried once, the best among the remaining set is selected.

Experiments show that the test application time depends on the order of application of the scan test vectors. We have experimented with some other alternative approaches which generate a different order of the scan test vectors. However the experimental results show that the approach presented here outperforms other heuristics.

#### V. EXPERIMENTAL RESULTS

We used the combinational part (available as ISCAS-89 SCAN circuits) of the ISCAS-89 benchmark circuits for test

---

**Algorithm 2:** Selecting Scan Test Vectors

---

**input** : Bipartite Graph  $G$  and  $C_{len}$   
**output**: A set of scan test vectors along with the circulation length for each scan test vector

- 1 **while** *Current fault coverage* < *desired coverage* **do**
- 2     Select maximum degree  $t_i \in TV$  ;
- 3     Perform sequential fault simulation as long as
- 4     no new fault is detected in consecutive  $C_{len}$  cycles;
- 5     *At the end of fault simulation*
- 6     Compute the circulation length for  $t_i$ ;
- 7     Drop the detected faults ;
- 8     **foreach** *detected fault*  $f_j$  **do**
- 9         Delete the all edges  $e_{kj}$  incident on  $f_j$  ;
- 10    **end**
- 11 **end**

---

pattern generation. Combinational ATPG tool ATALANTA is used to generate the test patterns for the experiment. The test patterns are obtained after enabling the compaction option of ATALANTA. We have used checkpoint theorem to generate the fault list for every circuit.

We varied the sequence length( $C_{len}$ ) for the circulation from 0 to 25 at a step size of 5. Table II shows the corresponding sequence length that maximizes the improvement for the same fault coverage provided by ATALANTA. The equation that we used for computing the improvement is:

$$Imp = \frac{(C_{FS} - C_P)}{C_{FS}} \times 100 \quad (1)$$

where  $C_{FS}$  is the number of clock cycles required in normal scan-based architecture using the test vectors generated by ATALANTA and  $C_P$  is the number of clock cycles required using the proposed methodology for the same circuit. In Table II, the different columns are as follows.

- *Init Vec* denotes the number of test-vectors generated by ATALANTA.
- *Scan Vec* represents the number of test-vectors applied in scan mode in the proposed methodology.
- *Circ Resp* denotes the number of vectors that are applied in circulating response mode.
- *Seq Len* is the value of  $C_{len}$  that maximizes the improvement.
- *Imp* is the improvement in terms of test application time, following Equation 1.
- *CPU Time* reports the time required for the test vector generation by a machine having Intel dual core cpu (1867 Mhz) with 2 GB RAM.

## VI. CONCLUSION

A new hybrid test architecture is described here and empirically evaluated, which shows significant improvement in test application time. The proposed method makes a good use of both scan-based and non-scan testing. Although the method is studied for the cases of serial scan chain, it can be extended

Circuit	#Init Vec	#Scan Vec	#Circ Resp	Seq Len	Imp(%)	CPU Time(S)
s1196	144	16	515	15	35.00	0.73
s1238	149	75	652	15	26.00	1.14
s1423	69	36	267	15	42.00	0.59
s1494	129	80	104	5	26.00	0.58
s5378	263	205	325	10	21.00	5.36
s9234.1	314	225	1386	20	26.00	31.8
s13207.1	466	350	1183	20	24.00	51.37
s15850.1	448	278	2491	25	36.00	74.98
s35932	69	18	139	15	72.00	18.66
s37417	902	615	3689	20	31.00	311.15
s38584.1	656	350	5030	25	46.00	206.31

TABLE II  
IMPROVEMENT WITH OPTIMAL SEQUENCE LENGTH

easily to multiple scan chains, Illinois scan, or other tree-based scan architectures.

As the test controlling scheme is very simple, this test architecture can be implemented for BIST very easily. The proposed method will be useful in transition testing, as circulating tests, being applied in non-scan mode, can be fed at-speed, and the corresponding errors can be accumulated and observed in the scanout mode. Further, the scheme helps in reducing test and response data, because, the number of vectors to be scanned in from the tester, as well as the response vectors to be scanned out are reduced significantly.

We believe that the overhead in terms of the hardware is comparable to the overhead that is typically incurred for BIST architecture. Also this extra hardware does not have any significant effect on the power profile of the circuit.

## REFERENCES

- [1] J. Rajski, J. Tyszer, M. Kassab, N. Mukherjee, R. Thompson, K. Tsai, A. Hertwig, N. Tamarapalli, G. Mrugalski, G. Eider, and J. Qian, "Embedded deterministic test for low cost manufacturing test", in Int. Test Conf., Baltimore, MD, pp. 301-310, Oct. 2002.
- [2] I. Pomeranz, L. N. Reddy, and S. M. Reddy, "COMPACTEST: A method to generate compact test sets for combinational circuits", IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 12, no. 7, pp. 1040-1049, Jul. 1993.
- [3] I. Hamazaoglu and J. H. Patel, "Test set compaction algorithms for combinational circuits", IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 19, no. 8, pp. 957-963, Aug. 2000.
- [4] S. Y. Lee and K. K. Saluja, "Test application time reduction for sequential circuits with scan", IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 14, no. 9, pp. 1128-1140, Sep. 1995.
- [5] I. Pomeranz and S. M. Reddy, "Test compaction for at-speed testing of scan circuits based on nonscan test sequences and removal of transfer sequences", IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 21, no. 6, pp. 706-714, Jun. 2002.
- [6] I. Pomeranz and S. M. Reddy, "Static test compaction for full-scan circuits based on combinational test sets and non-scan sequential test sequences", in 16th Int. Conf. VLSI Design, New Delhi, India, pp. 335-340, Jan. 2003.
- [7] W. Lai, C. Kung, and C. Lin, "Test time reduction in scan designed circuits", in European Design Automation Conf., pp. 489-493, Mar. 1993.
- [8] Y. Cho, I. Pomeranz, and S. M. Reddy, "On Reducing Test Application Time for Scan Circuits Using Limited Scan Operations and Transfer Sequences", IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 24, no. 10, Oct. 2005.
- [9] B. B. Bhattacharya, S. C. Seth, and S. Zhang, "Low-Energy BIST Design for Scan-based Logic Circuits" in International Conference on VLSI Design 2003.
- [10] B. B. Bhattacharya, S. C. Seth, and S. Zhang, "On Finding Consecutive Test Vectors in a Random Sequence for Energy-Aware BIST Design" in International Conference on VLSI Design 2005.
- [11] J. Savir and S. Patil, "On Broad-Side Delay Test", VLSI Test Symp., pp. 284-290, Sept. 1994.
- [12] I. Pomeranz and S. M. Reddy, "Autoscan: A Scan Design Without External Scan Inputs or Outputs", IEEE Trans. VLSI Syst., vol. 13, no. 9, Sept. 2005.
- [13] I. Pomeranz and S. M. Reddy, "Autoscan-Invert: An Improved Scan Design without External Scan Inputs or Outputs", IEEE VLSI Test Symposium (VTS 2007), pp. 416-421, May 2007.