

Design and Implementation of Online BIST Architecture Using SRAM Cells

Sheik Husseni

PG Scholar,

Department of ECE(VLSI),

Sree Vahini Institute Of Science & Technology.

Nadakuduru Dharmachari

Assistant Professor,

Department of ECE,

Sree Vahini Institute Of Science & Technology.

Abstract:

Input vector monitoring concurrent built-in self test (BIST) schemes perform testing during the normal operation of the circuit without imposing a need to set the circuit offline to perform the test. These schemes are evaluated based on the hardware overhead and the concurrent test latency (CTL), i.e., the time required for the test to complete, whereas the circuit operates normally. In this brief, we present a novel input vector monitoring concurrent BIST scheme, which is based on the idea of monitoring a set (called window) of vectors reaching the circuit inputs during normal operation, and the use of a static-RAM-like structure to store the relative locations of the vectors that reach the circuit inputs in the examined window; the proposed scheme is shown to perform significantly better than previously proposed schemes with respect to the hardware overhead and CTL tradeoff. As an extension, in the CUT we are using a 8*8 multiplier. We are using Xilinx version to verify the output. The performance analysis of the 8*8 multiplier is verified using BIST.

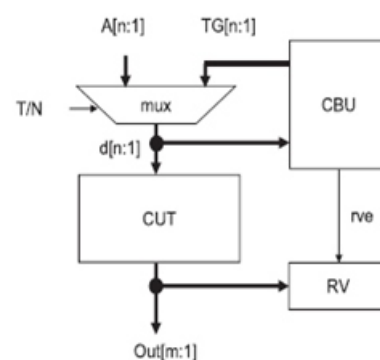
Keywords:

Built-in self-test (BIST), Circuit Under Test (CUT), Response verifier (RV), testing.

I. INTRODUCTION:

With the advance of VLSI technology, the capacity and density of memories is rapidly growing. The yield improvement and testing issues have become the most critical challenges for memory manufacturing. The BIST is used to detect and locate faulty cells of circuit under test. Input vector monitoring concurrent BIST techniques have been proposed to avoid this performance degradation. These architectures test the CUT concurrently with its normal operation by exploiting input vectors appearing to the inputs of the CUT; if the incoming vector belongs to a set called active test set, the RV is enabled to capture the CUT response.

The block diagram of an input vector monitoring concurrent BIST architecture. BIST utilizes a Test Pattern Generator (TPG) to generate the test patterns that are applied to the inputs of the Circuit Under Test (CUT). In off-line BIST, the normal operation of the CUT is stalled in order to perform the test. Thus, if the CUT is of critical importance for the function of the circuit, the total circuit performance is degraded. To avoid such performance degradation, input vector monitoring concurrent BIST techniques have been proposed, that exploit input vectors arriving at the inputs of the CUT during normal operation. Linear Feedback Shift Registers (LFSRs) have been by far the most popular devices for pseudo-random test pattern generation in BIST schemes. They have the advantage of very low hardware overhead. However, for circuits with random pattern resistant faults, high fault coverage cannot be achieved within an acceptable test length.



Input vector monitoring concurrent BIST.

Fig 1: CBIST

A. CUT (Circuit Under Test):

The CUT has n inputs and m outputs and is tested exhaustively; hence, the test set size is $N = 2^n$. Let us consider a combinational CUT with n input lines, as shown in Fig. 2; hence the possible input vectors for this CUT are 2^n . The proposed scheme is based on the idea of monitoring a window of vectors,

hence the possible input vectors for this CUT are 2^n . The proposed scheme is based on the idea of monitoring a window of vectors, whose size is W , with $W = 2w$, where w is an integer number $w < n$. Every moment, the test vectors belonging to the window are monitored, and if a vector performs a hit, the RV is enabled.

B. CBU (Concurrent BIST Unit):

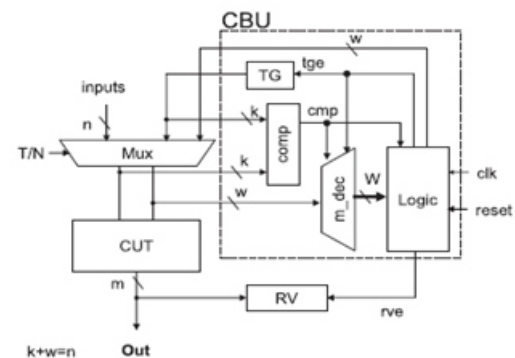
A typical BIST architecture that consists of Test Pattern Generator (TPG) usually implemented as a LFSR, Test Response Analyzer (TRA), Multiple Input Signature Register (MISR), CUT and BIST control unit. The approach uses the concept of reducing the transitions in the test pattern generated by conventional LFSR. The transition is reduced by increasing the correlation between the successive bits. The technique can operate in either normal or test mode, depending on the value of the signal labeled T/N. During normal mode, the vector that drives the inputs of the CUT (denoted by $d[n:1]$ in Fig. 1) is driven from the normal input vector ($A[n:1]$). A is also driven to a concurrent BIST unit (CBU), where it is compared with the active test set. If it is found that A matches one of the vectors in the active test set, we say that a hit has occurred. In this case, A is removed from the active test set and the signal response verifier enable (rve) is issued, to enable the m -stage RV to capture the CUT response to the input vector. C-BIST has low hardware overhead but very high concurrent test latency, since in every clock cycle the input vector is compared against only one active test vector. To drive down the concurrent test latency, four techniques have been proposed so far, namely Multiple Hardware Signature Analysis Technique (MHSAT, [3]), Order Independent Signature Analysis Technique (OISAT, [4]), windowed-Comparative Concurrent BIST (w -CBIST, [5]) and RAM-based concurrent BIST (RC-BIST [6]). These techniques accomplish to decrease the Concurrent Test Latency by increasing the number of active test vectors.

C. RV (Response Verifier):

When all input vectors have performed hit, the contents of RV are examined. During test mode, the inputs to the CUT are driven from the CBU outputs denoted $TG[n:1]$. The concurrent test latency (CTL) of an input vector monitoring scheme is the mean time (counted either in number of clock cycles or time units) required to complete the test while the CUT operates in normal mode.

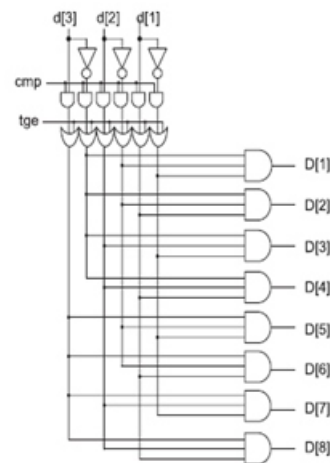
The Active test set Generator and Comparator of CBIST is a single Linear Feedback Shift Register (LFSR) and a comparator and thus the active test set consists of only one active test vector, which is the current value of the LFSR. During normal mode, the input vector is compared with the unique active test vector. If the two vectors match, the LFSR proceeds to the next state changing the active test vector and the Response Verifier is enabled.

II. Proposed Architecture:



Proposed architecture.

Fig 2: Proposed Architecture



Modified decoder design used in the proposed architecture.

Fig 3: Decoder Architecture.

A. Architecture of Proposed Scheme:

The bits of the input vector are separated into two distinct sets comprising w and k bits, respectively, such that $w + k = n$. The k (high order) bits of the input vector show whether the input vector belongs to the window under consideration.

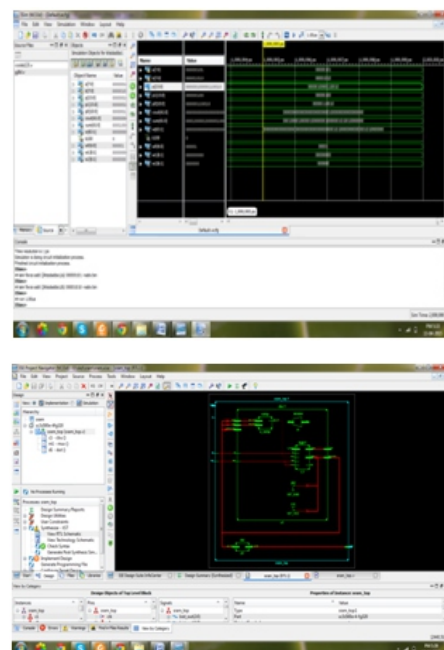
The remaining bits show the relative location of the incoming vector in the current window. If the incoming vector belongs to the current window and has not been received during the examination of the current window, we say that the vector has performed a hit and the RV is clocked to capture the CUT's response to the vector. When all vectors that belong to the current window have reached the CUT inputs, we proceed to examine the next window. The module implementing the idea is shown in Fig. 2. It operates in one out of two modes, normal, and test, depending on the value of the signal T/N. When T/N = 0 (normal mode) the inputs to the CUT are driven by the normal input vector. The inputs of the CUT are also driven to the CBU as follows: the k (high order) bits are driven to the inputs of a k-stage comparator; the other inputs of the comparator are driven by the outputs of a k-stage test generator TG. The proposed scheme uses a modified decoder (denoted as m_dec in Fig. 2) and a logic module based on a static-RAM (SRAM)-like cell, as will be explained shortly. The design of the m_dec module for w = 3 is shown in Fig. 3 and operates as follows. When test generator enable (tge) is enabled, all outputs of the decoder are equal to one. When comparator (cmp) is disabled (and tge is not enabled) all outputs are disabled. When tge is disabled and cmp is enabled, the module operates as a normal decoding structure.

B. DADDA MULTIPLIER:

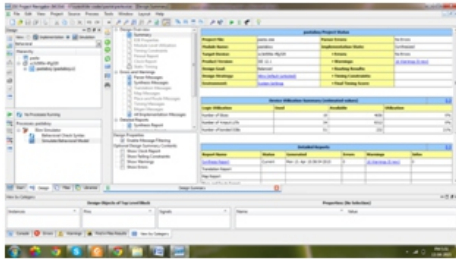
Dadda multipliers are the refinement of parallel multipliers first presented by Wallace in 1964. Dadda multiplier performs reduction whenever necessary. The maximum height of each stage is determined by final stage addition which consists of two rows of partial products. These two rows are then combined using a fast carry propagating adder (CPA). In the Wallace method, the partial products are reduced as soon as possible. In contrast, Dadda method does the minimum reduction necessary at each level to perform the reduction in the same number of levels as required by the Wallace method resulting in a design with fewer full adders and half adders. The disadvantage of Dadda method is that it requires a slightly wider, fast CPA and has a less regular structure than Wallace's. Figure 2 shows a 8x8 Dadda multiplier. Dadda [2] generalized and extended Wallace's results by noting that a full adder can be thought of as a circuit, which counts the number of ones in the input and outputs that number in 2-bit binary form. Using such a counter, Dadda postulated that, at each stage, only minimum amount of reduction should be

done in order to reduce the partial product matrix by a factor of 1.5. Dadda's method requires the same number of levels as that of Wallace method. However Dadda's method does the minimum reduction necessary at each level. This results in a design with fewer full adders and half adders. The number of (3,2) and (2,2) counters required is minimized in Dadda's technique compared to Wallace tree. The disadvantage of Dadda's method is that it requires a slightly wider fast Carry Propagate Adder (CPA) and has a less regular structure than Wallace. Fig 1 shows a 8x8 Dadda multiplier. The reduction process for a Dadda multiplier is developed using the following recursive algorithm. Wallace or Dadda algorithm for compressing partial products and 4x4 Dadda and Wallace tree. The proposed hybrid multiplier uses Wallace compression for the same groups and Dadda compression is remaining groups. Two groups are assigned Wallace compression and other two groups are assigned Wallace compression. Various hybrid multiplier combinations have been constructed and simulated.

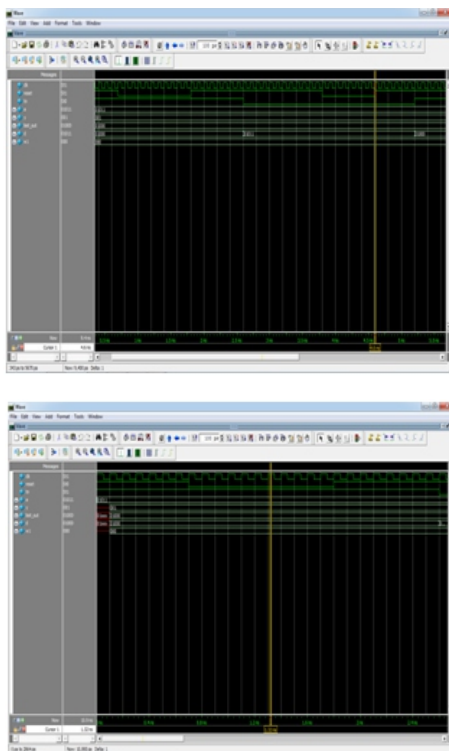
III. SIMULATION RESULTS: DADDA MULTIPLIER:



View rtl schematic:



Synthesis report:



IV. CONCLUSION:

In this brief, a novel input vector monitoring concurrent BIST scheme is proposed, which compares favorably to previously proposed schemes with respect to the hardware overhead/CTL tradeoff. BIST schemes constitute an attractive solution to the problem of testing VLSI devices. Input vector monitoring concurrent BIST schemes perform testing during the circuit normal operation without imposing a need to set the circuit offline to perform the test, therefore they can circumvent problems appearing in offline BIST techniques. The evaluation criteria for this class of schemes are the hardware overhead and the CTL, i.e., the time required for the test to complete, while the circuit operates normally. In this brief, a novel input vector monitoring concurrent BIST architecture has been presented, based on the use of a SRAM-cell

like structure for storing the information of whether an input vector has appeared or not during normal operation. The proposed scheme is shown to be more efficient than previously proposed input vector monitoring concurrent BIST techniques in terms of hardware overhead and CTL.

References :

- [1] M. Lin, Ed., "1997 Semiconductor Industry Annual Report," (in Chinese), Industrial Technology Research Institute (ITRI), Hsinchu, Taiwan, ITIS project report, 1997.
- [2] I. Voyiatzis, A. Paschalis, D. Gizopoulos, N. Kranitis, and C. Halatsis, "A concurrent BIST architecture based on a self-testing RAM," *IEEE Trans. Rel.*, vol. 54, no. 1, pp. 69–78, Mar. 2005.
- [3] I. Voyiatzis and C. Halatsis, "A low-cost concurrent BIST scheme for increased dependability," *IEEE Trans. Dependable Secure Comput.*, vol. 2, no. 2, pp. 150–156, Apr. 2005.
- [4] K. K. Saluja, R. Sharma, and C. R. Kime, "A concurrent testing technique for digital circuits," *IEEE Trans. Comput. Aided Design Integr. Circuits Syst.*, vol. 7, no. 12, pp. 1250–1260, Dec. 1988.
- [5] R. Sharma and K. K. Saluja, "Theory, analysis and implementation of an on-line BIST technique," *VLSI Design*, vol. 1, no. 1, pp. 9–22, 1993.
- [6] K. K. Saluja, R. Sharma, and C. R. Kime, "Concurrent comparative built-in testing of digital circuits," Dept. Electr. Comput. Eng., Univ. Wisconsin, Madison, WI, USA, Tech. Rep. ECE-8711, 1986.
- [7] I. Voyiatzis, T. Haniotakis, C. Efstathiou, and H. Antonopoulou, "A concurrent BIST architecture based on monitoring square windows," in *Proc. 5th Int. Conf. DTIS*, Mar. 2010, pp. 1–6.
- [8] M. A. Kochte, C. Zoellin, and H.-J. Wunderlich, "Concurrent self-test with partially specified patterns for low test latency and overhead," in *Proc. 14th Eur. Test Symp.*, May 2009, pp. 53–58.

- [9] S. Almkhaizim and Y. Makris, "Concurrent error detection methods for asynchronous burst mode machines," *IEEE Trans. Comput.*, vol. 56, no. 6, pp. 785–798, Jun. 2007.
- [10] S. Almkhaizim, P. Drineas, and Y. Makris, "Entropy-driven parity tree selection for low-cost concurrent error detection," *IEEE Trans. Comput. Aided Design Integr. Circuits Syst.*, vol. 25, no. 8, pp. 1547–1554, Aug. 2006.
- [11] C.-T. Huang, C.-F. Wu, J.-F. Li, and C.-W. Wu, "Built-in redundancy analysis for memory yield improvement," *IEEE Trans. Rel.*, vol. 52, no. 4, pp. 386–399, Dec. 2003.
- [12] J.-F. Li, J.-C. Yeh, R.-F. Huang, and C.-W. Wu, "A built-in self-repair design for RAMs with 2-D redundancies," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 13, no. 6, pp. 742–745, Jun. 2005.
- [13] C.-D. Huang, J.-F. Li, and T.-W. Tseng, "ProTaR: An infrastructure IP for repairing RAMs in SOCs," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 15, no. 10, pp. 1135–1143, Oct. 2007.
- [14] T.-W. Tseng and J.-F. Li, "A shared parallel built-in self-repair scheme for random access memories in SOCs," in *Proc. Int. Test Conf. (ITC)*, Santa Clara, CA, USA, Oct. 2008, pp. 1–9.
- [15] S.-K. Lu, C.-L. Yang, Y.-C. Hsiao, and C.-W. Wu, "Efficient BISR techniques for embedded memories considering cluster faults," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, no. 2, pp. 184–193, Feb. 2010.
- [16] L. R. Huang, J. Y. Jou, and S. Y. Kuo, "Gauss-elimination based generation of multiple seed-polynomial pairs for LFSR," *IEEE Trans. Comput. Aided Design Integr. Circuits Syst.*, vol. 16, no. 9, pp. 1015–1024, Sep. 1997.
- [17] Y. Zorian and A. Ivanov, "An effective BIST scheme for ROM's," *IEEE Trans. Comput.*, vol. 41, no. 5, pp. 646–653, May 1992.