

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES A NOVAL BISR APPROACH FOR EMBEDDED MEMORY SELF REPAIR

G. Sathesh Kumar^{*1} & V. Saminadan²

^{*1}Research Scholar, Department of ECE, Pondcherry Engineering College, Puducherry, India

²Professor, Department of ECE, Pondicherry Engineering College, Puducherry, India

ABSTRACT

As the density of embedded memory increases, manufacturing yields of integrated circuits can reach unacceptable limits. Normal memory testing operations require BIST to effectively deal with problems such as limited access and "at speed" testing. Built-in self-repair (BISR) techniques are widely used for the repair of embedded memories. One of the key components of a BISR circuit is the built-in redundancy-analysis (BIRA) module, which allocates redundancies according to the designed redundancy analysis algorithm. This project proposes a BIRA scheme for RAMs, which can provide the optimal repair rate using very low area cost and single test run of multiple single input change (MSIC) vectors in a pattern. Furthermore, the manifested errors are detected at the modules' outputs using novel voting, while the latent faults are detected by comparing the internal states of the memory modules. Upon detection of any mismatch, the faulty modules are located and the state of a fault-free module is copied into the faulty modules.

Keywords: BISR, RAM, MSIC, BIST, Memory Testing, Redundancy Analysis.

I. INTRODUCTION

A BISR scheme for RAMs with 2-D redundancy, i.e., spare rows and spare columns/IO, typically consists of a Built-in Self-Test (BIST) circuit, a Built-in Redundancy Analyzer (BIRA), and a reconfiguration mechanism. The BIST is used to detect and locate faulty cells of the RAM under test. There configuration mechanism is used to swap the defective element with the spare element. The BIRA is used for optimizing the allocation of the 2-D redundancy. The BIRA can collect fault information during the test process and allocate the available redundancies on the fly. Therefore, the BIRA has heavy impact on the repair efficiency of the BISR scheme. The complexity of the redundancy allocation for RAM with 2-D redundancy is an non-deterministic polynomial-time-hard problem. However, the number of redundancies of a RAM typically are small. Thus, it is still feasible to allocate redundancies using an exhaustive search algorithm. Many BIRA schemes for RAMs with 2-D redundancy have been developed. They can roughly be divided into two classes: optimal and non-optimal redundancy analysis (RA) schemes. An optimal RA scheme can find a pair solution if the repair solution exists. For example, the BIRA schemes uses different algorithms to achieve an optimal repair rate (RR). However, an optimal RA scheme needs either high area cost or long test time. A heuristic RA scheme attempts to obtain a compromise between the RR and the cost of area and test time. For example, the BIRA schemes reported in [9], [11] and use heuristic RA algorithms to allocate redundancies. A BIRA circuit realizing an optimal or a heuristic RA algorithm typically needs a cache like storage element to store the positions of faulty cells. Since only the positions of faulty cells in a RAM should be stored, a cache-like storage element can be considered as a compressed device image memory.

The advancement in IC technology increases integration of memories. As SOC size is shrinking, the major area on SOC is occupied by embedded memories. Thus memories in chip will decide the yield of the SOC. Increase in yield of memories in turn increase the yield of SOC. In [1], SOC yield increases from 2% to 10% by improving the memory yield from 5% to 10%. The techniques used for yield improvements in memories are Built-In-Self Test (BIST) and BISR. Many algorithms are proposed for spare allocation for defected memories. There dundancy is of 1D (only spare row or column) [3-4] or 2D (spare row and spare column), Hardware redundancy technique is one

effective technique for enhancing the yield of RAM. Memory built-in self-repair (BISR) techniques have been widely used to execute the fault detection, fault location, and redundancy allocation for RAMs with redundancies in a Soc.

The repair rate and area cost of the Built In Redundancy Analysis (BIRA) is mainly depends on the redundancy organization. The redundancy organization memory is divided into various segments. In which spare row and columns are used differently. Spare rows are used to replace entire row in the memory and the columns are divided in several spare column groups. Here the access time and area cost is induced due to additional multiplexers.

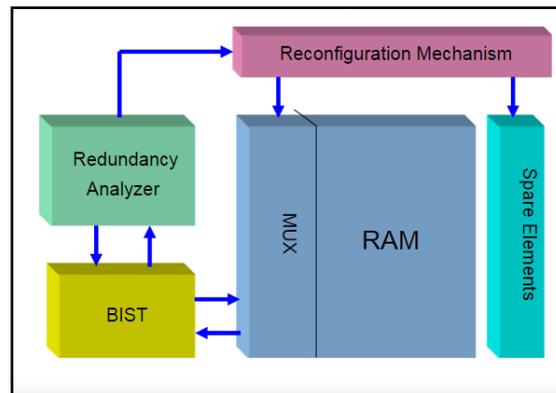


Fig 1: Built-in self repair for RAM

However different redundancy organization will lead to different area cost and repair rate. Since most of the memory faults are single cells, spare words are very efficient in reducing area. In this, the spare rows and spare columns are virtually divided into spare row blocks and spare column group blocks. SOC contains different RAMs with different address and data widths. If every RAM in SOC contains different repair circuit, area of the repair circuit increases. In order to reduce the area cost of repair circuit, reconfigurable repair circuit is used in which a single repair circuit is used for repairing multiple memories with different size and redundancies.

II. RELATED WORK

Rudrajit Dutta et al[1] is proposed here, an extremely low cost method to exploit these unused spare columns to improve the reliability of the memory by enhancing its existing error correcting code (ECC). Memories are generally protected with single-error-correcting, double-error-detecting (SEC-DED) codes using the minimum number of check bits.

Avijit Dutta et al[2] is proposed here for deriving an error correcting code that can correct all single errors and correct the most likely double bit errors i.e., double adjacent errors in a memory while completely eliminating the misscorrection of the most likely nonadjacent double errors.

Samuel Evain et al [3] proposed a way to increase the capacity of masking memory columns with isolated defective storage cells using spare memory columns. For this purpose, single error correction and double error detection (SEC-DED) codes already available for the protection against soft errors are extended such that all double-bit errors which affect a fixed sub-set of bit positions in the code words can be corrected. The cardinality of this sub-set is significantly higher than the number of spare columns. UmairIshaq et al [4] is proposed in this paper to efficiently fill the extra rows of the H-matrix on the basis of similarity of logic between the other rows. Optimization of the whole H-matrix is accomplished through logic sharing within a feasible operating time resulting in the reduced area and delay overhead. Mark Anders et al[5] is proposed here a 128-entry \times 128b content addressable memory (CAM) design enables 145ps search operation in 1.0V, 32nm high-k metal-gate CMOS technology. A high-speed 16b wide dynamic AND match-line, combined with a fully static search-line and swapped XOR CAM cell simulations show a 49% reduction of search energy at iso-search delay of 145ps over an optimized high-

performance conventional NOR-type CAM design, enabling 1.07J/bit/search operation. Vincent C.Gaudet et al[6] is proposed here a self-timed overlapped search mechanism for high-throughput content-addressable memories (CAMs) with low search energy. Most mismatches can be found by searching the first few bits in a search word. Consequently, if a word circuit is divided into two sections that are sequentially searched, most match lines in the second section are unused. Vincent Gripon et al[7] proposed a low-power content-addressable memory (CAM) employing a new algorithm for associativity between the input tag and the corresponding address of the output data. The proposed architecture is based on a recently developed sparse clustered network using binary connections that on-average eliminates most of the parallel comparisons performed during a search. Therefore, the dynamic energy consumption of the proposed design is significantly lower compared with that of a conventional low-power CAM design. Albert Lee et al[8] is proposed here a many big-data (BD) processors reduce power consumption by employing ternary content-addressable-memory (TCAM) [1-2] with pre-stored signature patterns as filters to reduce the amount of data sent for processing in the following stage (i.e., wireless transmission). To further reduce standby power, BD-processors commonly use non-volatile memory (NVM) to back up the signature patterns of SRAM-based TCAM (STCAM) [3] during power interruptions or frequent-off operations.

T.Ungerer et al[9] is proposed here a new real-time scheduling technique, called guaranteed percentage (GP) scheme that assigns each thread a specific percentage of the processor power. A hardware scheduler in conjunction with a multithreaded processor guarantees the execution of instructions of each thread according to their assigned percentages within a time interval of 100 processor cycles. E.Hadad et al[10] presents in this paper a lightweight CORBA fault-tolerance service called FTS. The service is based on standard portable features of CORBA, and in that respect is fully CORBA compliant, but does not follow the FT-CORBA specifications in areas where the authors felt the latter interfered with their other design goals. Luong D. Hung [11] is proposed here the technique to deal with the degraded resilience against soft errors. Only clean data can be stored in defective blocks of a cache. This constraint is enforced through selective write-through mechanism. An error occurring in a defective block can be detected and the correct data can be obtained from the lower level caches. T.C Hsia et al [12] is proposed here the redundancy to improve motion performance according to certain objective functions. The objective function can be either analytical or nonanalytical. For nonanalytical objective functions, a least-squares scheme is proposed to estimate the gradient vector. In addition, an approximation scheme is developed to compute the pseudo inverse of the Jacobian. Application of the scheme to a 4-link revolute planar robot manipulator is demonstrated through simulation.

III. SCHEME FOR STORING CHECK BITS IN UNUSED SPARE COLUMNS

This scheme can be implemented with very little modification to a normal memory that uses spare columns and is protected with an SEC-DED code. Fig. shows an example of the scheme assuming a single spare column. The additional logic that is added to support the scheme is the following:

- a. An extra XOR tree in the check bit generator and syndrome generator to support one additional check bit.
- b. An extra 2-input AND gate to disable the extra syndrome bit when determining error detection if the spare is used for repair.
- c. An extra 2-input OR gate in the correction logic for each data bit to disregard the extra syndrome bit if the spare is used for repair.

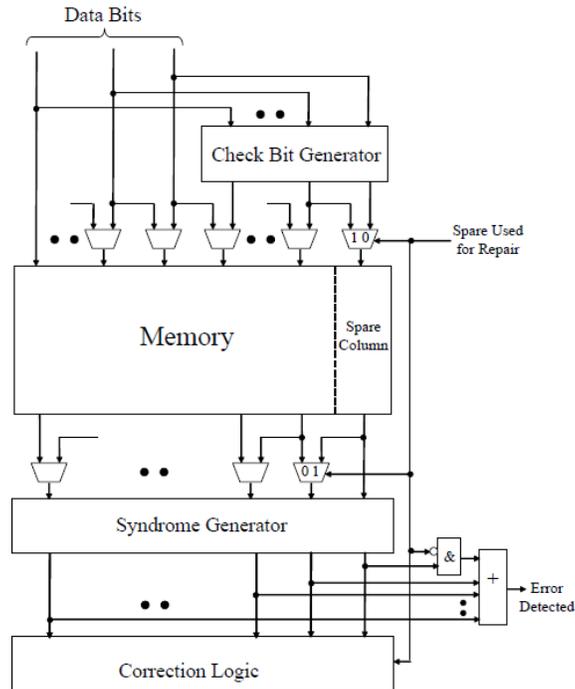


Fig1. Block Diagram for One Spare Column Utilization

If the spare is used for repair, then the MUXes at the input and output of the memory will shift the bits so that the defective column is bypassed. The control signal for the MUX on the far right will be a '1' if the spare is used for repair or if the spare column itself has a defect. If this control signal is a '0', then the spare is available for storing the extra check bit.

So if the spare is not used for repair, then the extra check bit generated by the check bit generator is stored in the spare column, otherwise, it is simply ignored. At the output of the memory, the extra syndrome bit that is generated is ignored if the spare is used for repair in which case error detection and correction are performed just as if that extra syndrome bit didn't exist. However, if the spare is not used for repair, then the extra syndrome bit is used to help increase the chance of detecting a multi-bit error as well as reduce the probability of miscorrection.

IV. PROPOSED METHODOLOGY

In our proposed method, an Optimized Built-In Self-Repair for Multiple Memories. A new built-in self-repair (BISR) scheme is proposed for multiple embedded memories to find optimum point of the performance of BISR for multiple embedded memories. All memories are concurrently tested by the small dedicated built-in self-test to figure out the faulty memories, the number of faults, and irreparability. After all memories are tested, only faulty memories are serially tested and repaired by the shared built-in redundancy analysis according to the sizes of memories in descending order. Thus, the fast test and repair are performed with low area overhead. To accomplish an optimal repair rate and a fast analysis speed, an exhaustive search for all combinations of spare rows and columns is proposed based on the optimized fault collection. The performance of the proposed BISR is located in the optimum point between the test and repair time, and the area overhead.

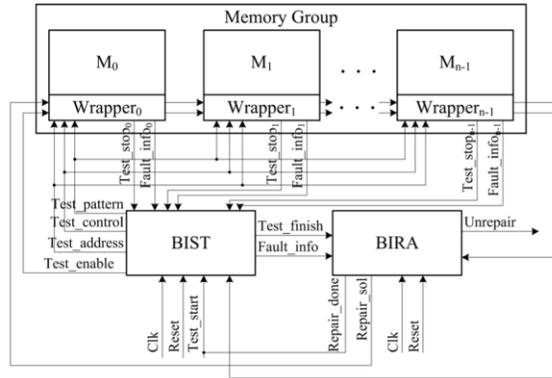


Fig2. Block diagram of the BISR architecture

Figure shows a block diagram of the proposed BISR architecture. It mainly consists of BIST and BIRA modules. The main purpose is to classify the memories as faulty or not, and the number of faults in each memory is stored in the dedicated wrapper. When the fault is detected by BIST, the fault information is sent to BIRA through the port Fault info. After the test is completed or stopped, the signal Test_finish is activated and BIRA executes the RA process to find repair solutions. In the repair procedure, RA based on the exhaustive search for all combinations of spare rows and columns is performed. If the memory under test cannot be repaired, the signal Unrepair is activated, and the test and repair are finished and the SoC is determined as a faulty chip due to the irreparable memory. If the memory under test can be repaired, the repair is performed by the repair solution. In this paper, a hard repair is performed using programmable electrical fuse (eFuse) schemes. The hard repair permanently repairs the faulty memory. Electrically programmable fuses are developed to perform hard repair using an eFuse and antifuse. After the repair is done, the signal Repair done is asserted. It is connected to the port Test start, and test is started for the next faulty memory.

A conceptual block diagram of the proposed BIST and wrapper modules. Assume that the number of memories is n . The BIST module consists of a test pattern generator (TPG), a test address generator (TAG), and a controller (CTR). The wrapper, which is dedicated to the memory, consists of a comparator (CMP) and a fault number register (FNR). When the signal Test start is asserted, TAG and TPG generate test patterns (Test pattern) and test addresses (Test address) for executing the adopted March test algorithm. The overall test procedure is controlled by the signal Test control. The clock and reset signals are provided through Clk and Reset. The CMP compares the results from the memory and expected responses to detect faults.

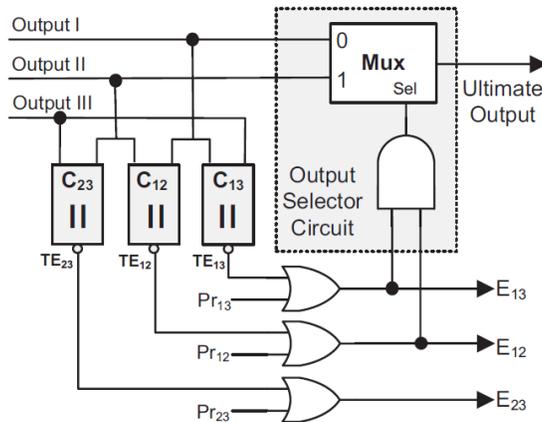


Fig3. Proposed voter

Additionally, the proposed voter can also detect possible faults occurring in the comparators. The architecture of the proposed voter is depicted in Fig. As shown in the figure, three comparators (C_{12} , C_{13} , and C_{23}) are used to represent any mismatch between TMR modules. As an example, $T E_{23}$ signal is activated once a mismatch between Outputs II and III is detected. If one of the modules generates an erroneous output (e.g., Output I), two of the comparators (here, C_{12} and C_{13}) will activate the mismatch signals (here, $T E_{12}$ and $T E_{13}$) and only one of the comparators (here, C_{23}) will not activate the corresponding mismatch signal (here, $T E_{23}$). In case of a faulty comparator (e.g., C_{13}), only the corresponding signal (here, $T E_{13}$) is activated and the other signals (here, $T E_{12}$ and $T E_{23}$) are deactivated.

According to the test order, the target memory under test is tested. Whenever a fault is detected, the fault information is sent through the port Fault_info target. When the detected number of faults reaches the number of faults stored in the first test, the test is stopped by the signal Test stop target. Then, the signal Test_ finish is activated to prevent test time from being wasted, while the test algorithm is being completed during the serial test. The test time is greatly reduced, because most faults are detected in the first few read operations of a March test.

A BISR technique for multiple embedded memories is proposed. To find optimum point of the performance of BISR for multiple embedded memories, the proposed BISR scheme is proposed. All memories are concurrently tested by the small dedicated BIST to figure out the faulty, the number of faults, and irreparability. After all memories are tested, only faulty memories are serially tested and repaired by the global BIRA according to the sizes of memories in descending order. The proposed BISR scheme finds the optimum point between the test and repair time, and the area overhead by maintaining the optimal repair rate. In addition, the verification procedure is simply conducted through the parallel test. Therefore, the proposed BISR scheme is a solution that trades off test and repair time, and area overhead to accomplish an optimal repair rate for multiple embedded memories in the SoC.

V. EXPERIMENTAL RESULTS

The proposed circuit is simulated and synthesized by using modelsim and xilinx12.1 which occurs low area than the existing. The experimental results are given in Table 1. The area of slice registers is taken for comparison, which is 11] in the case of existing method and it is only 7 in the proposed method. The area of IOB bonds in the proposed system consumes 20, which is comparatively less than conventional method. The area of LUT in the proposed system consumes 6, which is comparatively less than existing method. The evaluated implementation results of SLICE, LUT, IOB are listed and compared for both approaches which are tabulated in Table 1, and the corresponding bar chart comparisons are depicted Fig.4

Table 1. Performance parameters

S. No	Parameter	Existing	proposed
1	SLICE	11	7
2	LUT	19	6
3	IOB	26	20

VI. PERFORMANCE ANALYSIS

The Figure given below is shown that there is a considerable reduction based on no of transistors and the performance chart has been shown below in fig.

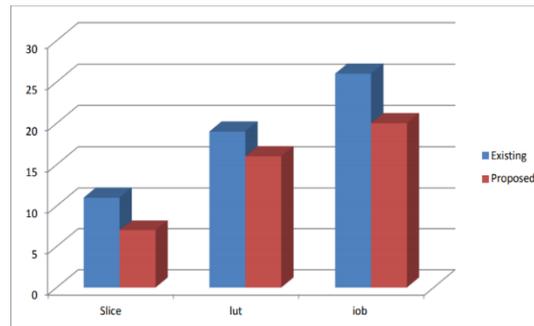


Fig4.Performance analysis

VII. CONCLUSION

A BISR technique for multiple embedded memories is proposed. To find permanent faults in memories SIC pattern as been applied to multiple memories. All memories are concurrently tested by the small dedicated BIST to figure out the faulty, the number of faults, and irreparability. The proposed BISR scheme finds the optimum point between the test and repair time, and the area overhead by maintaining the optimal repair rate. In addition, the verification procedure is simply conducted through the parallel test. Therefore, the proposed BISR scheme is a solution that trades off test and repair time, and area overhead to accomplish an optimal repair rate for multiple embedded memories in the SoC.

REFERENCES

1. Rudrajit Dutta, *Exploiting Unused Spare Columns to Improve Memory ECC*, 2009 27th IEEE VLSI Test Symposium.
2. Avijit Dutta, *Low cost adjacent double error correcting code with complete elimination of miscorrection within a dispersion window for Multiple Bit Upset tolerant memory*, 2012 IEEE/IFIP 20th International Conference on VLSI and System-on-Chip (VLSI-SoC).
3. Samuel Evain, *Programmable extended SEC-DED codes for memory errors*, 29th VLSI Test Symposium.
4. UmairIshaq, *Efficient Use of Unused Spare Columns to Improve Memory Error Correcting Rate*, 2011 Asian Test Symposium.
5. Marh Andres, *A 128×128b high-speed wide-and match-line content addressable memory in 32nm CMOS*, 2011 Proceedings of the ESSCIRC (ESSCIRC).
6. Vincent C.Gaudet, *High-Throughput Low-Energy Content-Addressable Memory Based on Self-Timed Overlapped Search Mechanism*, 2012 IEEE 18th International Symposium on Asynchronous Circuits and Systems.
7. Vincent Gripon, *Algorithm and Architecture for a Low-Power Content-Addressable Memory Based on Sparse Clustered Networks*, IEEE Transactions on Very Large Scale Integration (VLSI) Systems.
8. Albert Lee, *17.5 A 3T1R non-volatile TCAM using MLC ReRAM with Sub-Ins search time*, 2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers.
9. T. Ungerer , *A scheduling technique providing a strict isolation of real-time threads*, Proceedings of the Seventh IEEE International Workshop on Object-Oriented Real-Time Dependable Systems. (WORDS 2002).
10. E.Hadad, *FTS: a high-performance CORBA fault-tolerance service*, Proceedings of the Seventh IEEE International Workshop on Object-Oriented Real-Time Dependable Systems. (WORDS 2002).
11. Luong D. Hung *Utilization of SECDED for Soft Error and Variation-Induced Defect Tolerance in Caches*, 2007 Design, Automation & Test in Europe Conference & Exhibition.
12. T.C Hsia, *Joint trajectory generation for redundant robots*, Proceedings, 1989 International Conference on Robotics and Automation.
13. S.S Yau, *Schedulability in model-based software development for distributed real-time systems*, Proceedings of the Seventh IEEE International Workshop on Object-Oriented Real-Time Dependable Systems. (WORDS 2002).

14. G.Y Tang, *Compression of 3D objects with multistage colour-depth panoramic maps*, *Proceedings DCC 2002. Data Compression Conference.*
15. K.Pagiamtzis, *Content-addressable memory (CAM) circuits and architectures: a tutorial and survey*, *IEEE Journal of Solid-State Circuits.*