Fault Tolerant Design for Magnetic Memories

Arun Kumar P., P. Pandian and J. Raja Paul Perinbam
SENSE School,
SAS, VIT University, Vellore-632014, India
E.C.E Department, KCG College of Technology, Chennai, India

Abstract: This study presents a Fault Tolerant memory cores based on the property of Component Reusability, a method for Fault Tolerance for content addressable memories. The memories used in the design are 256, 512, 1024 and 2048 bytes. The fault is injected into the circuitry operation by using Automatic Test Pattern Generators (ATPGs). The design has been implemented in Cadence 90 nm technology and tested with Fault Injection Circuits and ATPG effectiveness was found out to be 100% at a frequency of 500 MHz.

Keywords: ATPG, BIST, fault tolerant, LFSR, magnetic memories

INTRODUCTION

Now-a-days in VLSI industry, testing has become a major criterion to determine the effectiveness of the product (Lei et al., 2008). Various testing methods have been proposed for BIST in order to find out the reliability of the design and also to identify the faults so that it can be corrected so that the design can be more effective (Mary and Stroud, 2009). Many researchers have proposed their own algorithm or modified the existing algorithm to find the effectiveness of fault coverage in various systems. Normally testing involves covering 100% fault coverage but in practice, we can detect only 95 to 99.99% of the total fault in a given circuit if the design has been made more efficient with respect to various parameters and the Total Fault Coverage (TFC) i.e., ATPG Effectiveness (Ian, 2007) can be calculated by the formula given below:

\[
\text{TFC} = \frac{\text{Number of faults detected}}{\text{Total number of faults considered}}
\]

Various Computer Aided Test (CAT) tools have been developed by various semiconductor companies for testing the reliability of the design so that we can detect more number of faults based on our design and we can able to cover a maximum percentage of fault coverage and make the system more effective. The basic idea (Zhiquan et al., 2009) of BIST is to program a particular block as Test Pattern Generator (TPG) and to test its embedded blocks that are configured as Circuit Under Test (CUT). BIST mainly relies on TPGs which are configured by using Linear Feedback Shift Registers (LFSRs). The advancement of BIST in ICs has led to various developments in various testing such

Magnetic Random Access Memory (MRAM): As the technology scales down to Nanometer, various advancements in memory design have been made but they lack in non-volatility. As many computer systems use only SRAMs, the biggest problem is that it has to be re-programmed at each power-up (Maddu and Oommen, 2010). By taking into the consideration of area and non-volatility, MRAM has more advantages as because it has data retention of 10 years and symmetric read-write of 25-35 ns. In this memory, the data is stored in two ferromagnetic plates separated by an insulator and the direction of its magnetization indicates the storage elements “0” and “1” as shown below in Fig. 1 (Su et al., 2006).

As seen from the Fig. 1, when the current is in parallel state, “0” is written in the cell and when it is in anti-parallel state, which offers high resistance, “1” is...
written. The read operation also takes the same methodology of high and low resistances. The digital model of MRAM has been designed using Verilog by taking into its consideration of data storage during ON/OFF power supply. Here the data is stored in a temporary register as shown in the Verilog Code which is given by:

\[
\text{if (!power) begin data out = data in; temp = data in;}
\]

where data in is the input to be stored and data out is the output and temp is the temporary register for storing the data in MRAM. The block diagram of MRAM is shown below in Fig. 2 which is modeled in Verilog.

In the above block diagram, “Address” is the address generator of the memory core which is obtained from the decoder, “Data In” is the input data of the memory, Read/Write is used for reading and writing the data in the memory core based on the condition low and high respectively and Power and clk are the input parameters of the memory core and “Data Out” is the output of the memory core. The MRAM array core for 128 bits, 512 bits, 1 K and 2 K are based on the table shown below in Table 1.

<table>
<thead>
<tr>
<th>Memory size</th>
<th>Address</th>
<th>Data in</th>
</tr>
</thead>
<tbody>
<tr>
<td>128 bits</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>512 bits</td>
<td>32</td>
<td>16</td>
</tr>
<tr>
<td>1 K</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>2 K</td>
<td>64</td>
<td>32</td>
</tr>
</tbody>
</table>

The array diagram of MRAM Core (MC) is shown in Fig. 3. Here the decoder is enabled with enable input. When it is “LOW”, all outputs of the decoder are “0” and none of the memory words are selected. When it is high, one of the words are selected as determined by the decoder and the read/write input determined the operation of the memory cell when the power is in ON state. During the write operation, the data available in the input lines are transferred into the MCs of the selected words. During the read operation, the bits of the selected words are transferred into the OR gate which is the output terminal. When the power gets switched OFF, the values stored in the temporary register are transferred to the data out register thus maintain no loss in data.

**Table 1: Memory array size**

**Fig. 2: Block diagram of MRAM core**

**Fig. 3: Array diagram of magnetic memory core**

**Fig. 4: Block diagram of MC testing**

**METHODOLOGY**

**Low power ATPG design:** The basic idea in low power ATPG is reducing the number of test patterns and at the same time reducing the switching activity at CUT (Ravi et al., 2007). The switching activity can be controlled by filling don’t care conditions (X Bits) which will try to reduce the leakage power and also reduce the test length (Kumar et al., 2012). The block diagram for our test scheme is shown below in Fig. 4. The main objective of the ATPG in memory testing is that to set the required state of the circuit and also to detect the faulty node of the circuit which is used to separate the fault free conditions when the test patterns are being generated. The Pseudo-random TPG are mainly generated by the LFSR which is fed into the memory device which acts as CUT and its output is fed into MISR which acts as Signature Analyzer.

The following objectives were taken into consideration for Low Power analysis:

- The switching activity of the circuit where only one half of the inputs are fed into the CUT where the number of transitions on the inputs are reduced.
Fault tolerant design: As the size of the memory increases in various embedded streams, effective test algorithms to be applied to isolate the faulty ones and should be repaired by using spare elements. The faults in the design are mainly caused by system specification, materials used in the manufacturing and operating parameters. The user has to develop a reduced functional data set which contains the Fault Models which will make easy to locate the fault (Van De Goor and Verruijt, 1990). To model these defects, we use the abstract models to develop the test stimuli and rectify these faults. As the MRAM is made of Ferro Magnets and the application of high currents which allows to write and read 1 and 0 so we consider the faults in terms of operating parameters which causes coupling faults and we develop a logic model to identify those defects. Consider the array diagram for 128 bit MRAM in Fig. 5 showing the coupling faults which are indicated in RED.

The output of each MC is connected to XNOR Gate which indicates the coupling faults i.e., if both the MC has same data to be read or written then a coupling fault has occurred. In this case, we have to isolate the faulty cell and provide a technique which will replace the faulty cell to work properly. The MARCH algorithm proposed in the previous technique for embedded memories in terms of coupling faults occupies up to 70 to 90% of the chip area and does not provide 100% Fault Coverage (Nor et al., 2012). The testing method provided should be more transparent and should cover all aspects of faults in the design. For this aspect we use Transparent BIST technique. Many algorithms have been proposed for Memory Testing and they are able to provide FC<100% and the repair rate of the BISR technique provides only 90-96% (Lu et al., 2012). We provide a technique of Component Reusability which never utilizes any algorithm for MBIST Technique rather it only provides a Fault Tolerant Routing which enables the Hardware to regain its position at the failure with less access time for repair and so in this case the FC = 100% with less area. Our design has an advantage of correction both in Static and Dynamic background as compared to both Single and Dual Port SRAMs (Nor et al., 2011) with less cycle Test Time. Consider the diagram below in Fig. 6 which shows the Reusability technique.

The Hardware Reusability concept works only when the system utilizes same hardware concepts. In our design, the memories are built with the same hardware resources where the functionality of each module is same and the outputs of each memories are connected to XNOR Gate which shows when two memories does the same job, a coupling Fault has occurred. Here in our diagram, consider MC2 has become faulty and so the output generates a Control signal “1” indicating a Fault has occurred in a particular location. The signal now switches to the nearby cell which will also act as a spare for the faulty one and so in this case of design we need not use a spare cell thus reducing the area. The major concept in our design takes into the consideration of Configurable Architecture which was introduced for various algorithms for Memory BIST (Atieh et al., 2011). As the faulty cell gets replaced by the dummy cell, we take the concept of structural hazard where two cells operate simultaneously which increases the power consumption. In order to reduce this activity, we introduce the concept of clustering (Ahmed and Abdallatif, 2011) which allows only one cell to be operated depending upon the read write operation. This was primarily utilized in SRAM where modified MARCH-C Algorithm was used for testing which takes only one operation at a time i.e., sequence. As seen in Fig. 6, when the control signal reaches high state, it must switch to MC1 which now acts as a dummy cell. The data’s of MC2 are stored in a temporary cell which are fed to MC1 during failure conditions and now
Fig. 7: BIST for MRAM

depending upon the ADDRESS Logic, data transfer occurs between the two memories where at a time only one cell is activated for both read write operations taking into the considerations of Power signal as show in Fig. 2. Thus we overcome the disadvantage of clustering employed in SRAM where we are able to achieve about 70% power reduction.

EXPERIMENTAL SETUP

We employ BIST architecture as shown below in Fig. 7. Rita and Gupta (2011) where the test functions are embedded in the core. The TPGs are given by LFSR to the CUT. The faults to the circuit are injected by the Fault Injection circuits as designed by the user called as differential faults. The operation takes place in parallel so that along with the test patterns, the faults are also being injected and the test response analyzer checks the value at each stage and reports the error. The error signal is sent back to the neighboring cell where both the memories are interconnected and it gives the signal to Non Faulty Cell thus replacing the faulty node by using the technique of Component Reusability.

The main objective in this design is we have to take care of routing at the shortest path which will reduce the interconnect delays. The routing has to be made manually by the user so that he can visualize the components having the same functionality for both combinational and sequential circuits. This technique provides a DFT support for non volatile RAMs (Andreas and Schmutz, 2011) which is automatic and provides the Test Patterns and injects the Faults. In order to operate the Test at a high frequency of 500 MHZ, we develop an on chip clock generator which works as a digital PLL by using the technique of Feed Forward Technique (Xin et al., 2010) to provide constant clocking so that it works synchronous with the system clock. During failure conditions of the CUT, i.e., whenever faults are being injected, the CUT must auto correct itself within a short span of time to the normal operation. For this aspect, we design a on chip clock generator which utilizes the function of LTG (Songwei et al., 2010) which is synchronous with the PLL which reduces the delay to 2 clock cycles for the system to regain its original point with a delay range of 200.56 ns which is for a short range of time with respect to the Reference signal so that no dithering effect takes place within the system. To overcome this problem, we develop a multiplexer based clock tree which will help the Reference signal to retain its value so that the system never shifts its original clock and the operating clock. In order to obtain low power consumption in the PLL, the REF clock should be kept at a much lower frequency than the system clock which generates the clock pulses for the BIST (Liangge et al., 2010) which utilizes TDC with the principle of pseudo inverter chain based architecture.

CONCLUSION

As MRAM is the emerging memory device which finds its applications in many computer related areas where electric supply is the main concern. As this memory is the future of all Non Volatile Memories in computer and embedded applications where data preservation is the major concept. As technology shrinks, the size of this memory also comes down when compared to other memory and also offers high programmability for both read and write. We have designed a Fault Tolerant MRAM taking into the consideration of attacks by the supply voltage as when high voltage drifts causes the device to operate inaccurately. The data has been modified into the digital model where the Test Patterns along with the Faults are being injected into the system and we have employed the component reusability technique which is also called as hardware reusability instead of employing MARCH Algorithm and the Test Coverage was obtained as 100%. The whole model has been designed in Verilog programming and synthesized in cadence. When compared to the previous implementation (Zhiquan et al., 2009), we are able to find that our rectification takes only 2 clock cycles to correct the fault and this method saves more hardware complexity instead of applying extra hardware or providing spare circuits or applying algorithms to correct the fault.

We have designed a Coupling Fault which covers only 2 cells and developed a hardware reusability which covers 100% Coverage. The Future development of this work can be extended to increasing the coupling faults to 3 or more cells and bring a short span clock cycle to correct the fault.

REFERENCES


