Survey of low power testing of VLSI circuits

P. Basker, A. Arulmurugan

Department Of Ece, Kongu Engineering College, Perundurai, Tamil Nadu, India

Email address: contactbasker@rediffmail.com (P.Basker), arul@kongu.ac.in (A.Arulmurugan)

To cite this article:

Abstract: The System-On-Chip (SoC) revolution challenges both design and test engineers, especially in the area of power dissipation. Generally, a circuit or system consumes more power in test mode than in normal mode. This extra power consumption can give rise to severe hazards in circuit reliability or, in some cases, can provoke instant circuit damage. Moreover, it can create problems such as increased product cost, difficulty in performance verification, reduced autonomy of portable systems, and decrease of overall yield. This paper surveys about the available low power testing techniques during testing. It also suggests some advantages and disadvantages associated with every techniques.

Keywords: Dft, Bist, Lfsr, Cut, Atpg

1. Introduction

Modern chip design has greatly advanced with recent silicon manufacturing technology improvements that escalate transistor counts, increasing a chip’s complexity while maintaining its size. This phenomenon will continue, resulting in at least 10 of today’s microprocessors fitting onto a single chip by 2005. Consequently design and test of complex digital circuits imposes extreme challenges to current tools and methodologies. VLSI circuit designers are excited by the prospect of addressing these challenges efficiently, but these challenges are becoming increasingly hard to overcome [1-3]

Test currently ranks among the most expensive and problematic aspects in a circuit design cycle, revealing the ceaseless need for innovative, test-related solutions. As a result, researchers have developed several techniques that enhance a design’s testability through DFT modifications and improve the test generation and application processes. Traditionally, test engineers evaluated these techniques according to various parameters: area overhead, fault coverage, test application time, test development effort, and so forth. But now, the recent development of complex, high-performance, low-power devices implemented in deep-submicron technologies creates a new class of more sophisticated electronic products, such as laptops, cellular telephones, audio- and video-based multimedia products, energy efficient desktops, and so forth. This new class of systems makes power management a critical parameter that test engineers cannot ignore during test development.

Test or DFT engineers find their main motivation in considering power consumption during test in a circuit’s consumption of more power in test mode than during normal operation. Zorian showed that the test power (the power consumed during test) could be twice as high as the power consumed during the normal mode. Several reasons cause this increased power usage[3]. First, test efficiency correlates with toggle rate; therefore, in the test mode the switching activity of all nodes is often several times higher than during normal operation. Second, test engineers use parallel testing in SoCs to reduce the test application time, which might result in excessive energy and power dissipation. Third, the DFT circuitry designed to reduce the test complexity is often idle during normal operation but might be intensively used in the test mode. Fourth, successive functional input vectors applied to a given circuit during system mode have a significant correlation. In contrast, the correlation between consecutive test patterns can be low. For example, in a signal-processing circuit for speech recognition, the input vectors behave predictably, with the least-significant bits more likely to change than the most-significant bits. Similarly, in high-speed circuits that process digital audio and video signals, the inputs to most of those modules change relatively slowly. In fact, designers of low-power circuits take advantage of this consistent behavior when they determine a circuit’s thermal and electrical limits, and system packaging requirements. In contrast, there is no definite correlation between the successive
test patterns generated by an automatic test-pattern generator for external testing or the patterns produced by a linear feedback shift register (LFSR) for built-in self test (BIST). This lack of correlation can result in significantly greater switching activity in the circuit during test than during normal operation. Because power dissipation in CMOS circuits is proportional to switching activity, test’s excessive switching activity can cause catastrophic problems, as detailed later. Although academic research on low-power design remains nearly independent of that for test, industrial practice requires ad hoc solutions for considering power consumption during test application. [7] Practiced solutions include over sizing power supply, package, and cooling to withstand the increased current during

testing (test engineers insert breaks into the test process to avoid hot spots); testing with reduced operating frequency; and system-under-test partitioning and appropriate test planning.

The first solution increases both hardware costs and test time. Although the second proposal uses less hardware, the reduced frequency increases test time and might lead to a loss of defect coverage because the reduced frequency can mask dynamic faults. Moreover, this solution reduces power consumption but lengthens test time, so it does not reduce the total energy consumed during test. The third solution of test partitioning and test planning detects dynamic faults, but increases hardware costs and test time. To provide an adequate response to these industrial needs, various researchers have proposed solutions for power problems encountered during test. [1-3]

2. Energy and Power Modeling

Power consumption in CMOS circuits can be static or dynamic. Leakage current or other current drawn continuously from the power supply causes static power dissipation. Dynamic dissipation occurs during output switching because of short-circuit current, and charging and discharging of load capacitance. For existing CMOS technology, dynamic power is the dominant source of power consumption, although this might change for future high-scale integration. The average energy consumed at node \( i \) per switching is \( \frac{1}{2}CV_{DD}^2 \), where \( C_i \) is the equivalent output capacitance, and \( V_{DD} \) is the power supply voltage. Therefore, a good approximation of the energy consumed in a period is \( \frac{1}{2}C_{PP}V_{DD}^2 \) where \( s_i \) is the number of switchings during the period. Nodes connected to more than one gate are nodes with higher parasitic capacitance. Based on this fact, as a first approximation we assume capacitance \( C_i \) to be proportional to the fan-out of node. Therefore, an estimation of the energy \( E_i \) consumed at node \( i \) during one clock period is

\[
E_i = \frac{1}{2} F_i C_p V_{DD}^2, \quad \text{where} \quad C_p = \text{the circuit’s minimum parasitic capacitance}. 
\]

According to this expression, estimating energy consumption at the logic level requires the calculation of fan-out \( F_i \).

3. Terminology

Test power is a possible major engineering problem in the future of SoC development. As both the SoC designs and the deep-submicron geometry become prevalent, larger designs, tighter timing constraints, higher operating frequencies, and lower applied voltages all affect the power consumption systems of silicon devices. [4]

3.1. Energy

The total switching activity generated during test application, energy affects the battery lifetime during power up or periodic self-test of battery-operated devices.

3.2. Average Power

Average power is the total distribution of power over a time period. The ratio of energy to test time gives the average power. Elevated average power increases the thermal load that must be vented away from the device under test to prevent structural damage (hot spots) to the silicon, bonding wires, or package.

3.3. Instantaneous Power

Instantaneous power is the value of power consumed at any given instant. Usually, it is defined as the power consumed right after the application of a synchronizing clock signal.

Elevated instantaneous power might overload the power distribution systems of the silicon or package, causing brown-out.

3.4. Peak Power

The highest power value at any given instant, peak power determines the component’s thermal and electrical limits and system packaging requirements. If peak power exceeds a certain limit, designers can no longer guarantee that the entire circuit will function correctly. In fact, the time window for defining peak power is related to the chip’s thermal capacity, and forcing this window to one clock period is sometimes just a simplifying assumption. For example, consider a circuit that has a peak power consumption during only one cycle but consumes power within the chip’s thermal capacity for all other cycles. In this case, the circuit is not damaged, because the energy consumed which corresponds to the peak power consumption times one cycle will not be enough to elevate the temperature over the chip’s thermal capacity limit (unless the peak power consumption is far higher than normal).

4. Problems Induced by Excessive Test Power

When dealing with high-density systems such as modern ASICs and SoCs, a nondestructive test must satisfy all the power constraints defined in the design phase. In addition
to preventing destruction of the CUT, cost, reliability, autonomy, performance-verification, and yield-related issues motivate power consumption minimization during test.[5] The cost constraints of consumer electronic products typically require plastic packages, which impose a tight limitation on power dissipation.

Unfortunately, excessive switching activity during test leads to increased current flows in the CUT, making the use of expensive packages for the removal of excessive heat imperative. Moreover, electro migration causes the erosion of conductors and subsequently leads to circuit failure. Because temperature and current density are major factors that determine electro migration rate, elevated temperature and current density (caused by the test’s excessive switching) severely decrease CUT reliability.

This phenomenon is even more severe in circuits equipped with BIST because such circuits might be tested frequently in, for example, online BIST strategies. Not only the reliability but also the autonomy of battery-powered remote and portable systems suffers from increased activity. Remote system operation occurs mostly in standby mode with almost no power consumption, interrupted by periodic self-tests. Hence, power savings during test mode directly prolong battery lifetime.

Two points emphasize the relevance of this power minimization problem. First, the current trend in circuit design toward circuit miniaturization (for portability, for example) prevents the use of special cooling equipment for removing excessive heat during test. Second, the growing use of at-speed testing for identifying slow chips no longer permits compensating for increased power dissipation by reducing test frequency. In the past, tests typically lower rates than a circuit’s normal clock rate.[2]

5. Methods

5.1. Low Transition TPGs

One common technique to reduce test power consumption is the design of low transition TPGs. Most of these techniques modify the design of the LFSR (or other forms of TPGs such as cellular automata) in such a way as to reduce the transitions in the primary inputs of the CUT for test-per-clock BIST or inside the scan-chain for scan-based BIST. An example of the low transition TPG for test-per-clock schemes is the approach presented in [5]. This approach, called DS-LFSR.

On the other hand, a TPG for low power consumption in scan-based BIST (test-per-scan BIST) is presented in [6]. The proposed design, called low transition random test pattern generator (LT-RTPG), is composed of an LFSR, a k-input AND gate, and a toggle flip-flop T-FF. Some cells of the LFSR are connected with the inputs of the k-input AND gate, and the output of the AND gate is connected with the CUT (the T-FF output will not toggle in ncells will have the same value in most cases.

Thus the power while scanning-in a test vector not while scanning out the captured response. Also, in order to get a high fault-coverage, a long test sequence is needed. Put of the T-FF, and the output of the T-FF is connected with the scan-chain input Sin). Since the output of the AND gate (input of the T-FF) is 0 in most of the cases, of the clock cycles, and hence the transition probability in the CUT will decrease. The main drawback of this system is that it reduces the average power while scanning-in a test vector not while scanning out the captured response. Also, in order to get a high fault-coverage, a long test sequence is needed.

5.2. Test Vectors Reordering

The test vectors reordering techniques aim to reduce the switching activity by modifying the order in which the test vector is applied to the CUT, thereby reducing test frequency. In the past, tests typically lower rates than a circuit’s normal clock rate.

5.3. Scan Cells Reordering Techniques

Another category of techniques used to reduce the power consumption in scan-based BIST is the use of scan-chain ordering techniques [9]. Changing the order of the scan cells in each scan-chain can reduce the switching activity, and hence power dissipation, in scan designs. In the case of a deterministic set of test patterns, the best order of cells is the one that gives the best compromise between reducing the transitions in the scan cells while scanning in test patterns and while scanning out captured responses.

Scan-chain ordering is one of the most popular algorithms to reduce the power consumption in scan-based BIST because this techniques of scan-chain ordering aim to reduce the average power consumption during scanning in
of test vectors and scanning out of captured responses by arranging the scan cells which cause more internal circuit transitions to the positions with low transition weights in the scan chain. As a simple example to demonstrate how cell ordering algorithms reduce the number of transitions in the scan-chain, assume that there is a scan-chain with 3 flip-flops: A, B, and C with initial values of “000”. Assume that the deterministic test vector 101 is to be scanned into the scan chain (i.e. ABC = 101). The scan shift with each clock. Since there are 3 flip-flops then 3 clock cycles are needed to scan in the test vector.

The arrows indicate that a transition has occurred in a particular cell in between two adjacent clock cycles. It is clear that the scan-chain with the new order ACB (i.e. the test pattern ACB = 110), this will reduce the number of transitions in the scan-chain to 2 .The advantage is it does not require additional hardware, it has no effect on fault coverage . However, its main drawback is routing congestion problems during scan routing. Also, most of the at scanning in this vector produces 6 transitions in the scan chain will be instead of 6.

![Fig 2. Effect of cell ordering in the number of transitions. (a) Original order and (b) Cells after reordering.](image)

Although these algorithms mainly aim to reduce average power consumption, they can reduce the peak power that may occur in the CUT during the scanning cycles, but not the capture power that may result during the test cycle (i.e. between launch and capture). However, some of the ordering algorithms aim to order the scan-chain cells in order to minimize the peak power that may arise during the test cycle (capture power) [10]. But these algorithms for reducing capture power are orthogonal with the algorithms that aim to reduce the scan power. None of the techniques in the surveyed literature tried to find a compromise between these algorithms or to merge the two techniques together in order to minimize the overall peak power.

### 5.4. X-Filling Techniques of Test Cubes

Test cubes are test vectors where the values of some bits are left unspecified as X’s (don’t care bits); for example 1X00X1 is a test cube. In scan-based BIST, a test vector that detects many targeted faults may contain many don’t care bits (X). In conventional scan ATPG, each X bit is filled with a 0 or 1 randomly since this will not affect the fault coverage. In fact, the number of X bits in each test cube is typically high [11]. The main aim of such techniques is to assign a value to the X bits in the test cube so that the number of the transitions in the scan cells is minimized, hence reducing the transitions in the scan-cells, which leads to a reduction in the overall switching activity in the CUT during shift cycles [12].

The non-random X-filling techniques can be divided into 3 approaches in order to reduce shift power: Minimum transition filling technique (MT-filling), also called adjacent filling, where the most recent non-X bit is used to fill successive X values until X bit is reached.

- **0-Filling.** Set 0 to all X-bits.
- **1-Filling.** Set 1 to all X-bits.

As a simple example, Consider the test cube [0XX1X0X1XX1X0X]. By applying the above filling techniques the resulting vectors and the number of transitions will be

- **MT-Filling:** 00011001111100 (4 transitions)
- **0-Filling:** 00010001001000 (6 transitions)
- **1-Filling:** 01111011111101 (5 transitions)

### 5.5. Vector Filtering Techniques

The test vectors that are generated by TPGs such as LFSRs are pseudorandom vectors. The fault detection capability of these vectors quickly reaches diminishing returns. Hence, after running a sequence of test vectors and detecting many faults, then only a few of the subsequent test vectors can still detect new faults.

The vectors that do not detect new faults, but do consume power when applied to the CUT, can be filtered or inhibited from being applied to the CUT [13]. These algorithms, in general, use extra logic (e.g. decoder circuitry). Using prior knowledge of the sequences of test vectors generated by TPGs such as the LFSR, they can prevent some sequences from being transmitted to the CUT by knowing the first and last vectors in this sequence. Thus they reduce the power consumption in the CUT.

### 5.6. Shift Control Technique
In scan-based BIST, during scan shifting of test vectors, the combinational part of the CUT will have many transitions although the test vector is not yet applied. There are many techniques that try to eliminate the switching activity in the combinational part of the CUT during scan shifting cycles [14]. This can be achieved by identifying a test vector, called input control or shift control, which is applied to the CUT’s primary inputs during scan, thereby minimizing or eliminating switching activity in the combinational part of the CUT. Then the then the scanning of a test vector will not affect the output of this gate; the output of the AND gate is an input to other parts of the CUT, and hence, a reduction to the switching activity in the CUT is achieved.

5.7. Selection of LFSR Parameters

Some parameters of the LFSR may affect the power consumption of the CUT, namely the seed and the characteristic polynomial of the LFSR. In [15] it is found that the WSA obtained for a given primitive polynomial of the LFSR strongly depends on the seed than to select the best polynomial. Furthermore, characteristic polynomial is not critical regarding the power consumption in the CUT, it is recommended in [16] to use the LFSR that needs the minimum number of XOR gates in its feedback since this will reduce area overhead and reduce the power consumption in the LFSR itself.

5.8. Low Power Test Vector Compaction

In scan-based circuits, in order to reduce the test data volume, compacting techniques are introduced to merge several test cubes. However, compacting test vectors greatly increases the power dissipation (it could be several times higher). Thus, low power test vector compaction techniques have been introduced to minimize the number of test cubes generated by the ATPG tool by merging test cubes that are compatible in all bit positions under a power constraint [17]. by carefully merging the test cubes in a specific manner, the number of transitions in the scan-chain can be minimized. Thus, a greedy heuristic procedure was used to merge test cubes in a way that m reductions in average and peak power consumption can be obtained by using this approach. As a simple example to demonstrate how compaction can significantly affect the power consumption, assume that there are three test cubes in a system: C = 0X0X0X, C = X1X1X1, and C = 0XXX11. By using the MT-filling technique described , the number of transitions in C is 0 (no transitions), in C is 0, and in C is 1. However, using this method, all the test vectors after MT-filling must be applied to the scan-based Minimizes the number of transitions. Instead, merging test cubes together is a good way to reduce test data and test application time. Thus, C1 can be merged with C2; since they are compatible in all bit positions, they C1, C2, and C3, the test cubes C2 and C3 can be merged with C2 since they are not compatible.

5.9. Scan Architecture Modification

This technique involves modifying the scan architecture by inserting new elements and partitioning the scan-chain into segments. In [18] the scan-chain is partitioned into N segments where only one segment is active at a time. This technique reduces the average power consumption in the CUT, but it will not affect the power will be enabled by using the gated clock trees instead of scan enable signals as was used in the previous technique. A similar approach can be used to reduce both peak and average power consumption as described in [19] where mutually exclusive scan segment activation signals are used in the scan architecture to get high reductions in peak power consumption. Other techniques based on scan architecture modifications are described in [20].

5.10. Scan Clock Splitting

In the scan-chain. The use of such a system reduces the average peak power consumption without increasing test time. This technique modifies the scan clock in order to reduce the power consumption during scan testing . In [21] a technique based on scan clock splitting is presented. It involves reducing the operating frequency of the scan cells during scan shifting without modifying the total test time. In this technique, two clocks have been used; each of them is at half the frequency of the system clock, and is used to activate half of the scan cells uses a staggered clock scheme to reduce the peak power consumption during scan testing is presented in[21].

5.11. LFSR Reseeding and Test Data Compression Technique

Considering together the problems of test data compression and low power test is very important. A long sequence of test vectors which also means high energy consumption since the total energy is a function of time. Hence, compression of test data leads to reduction to the total energy. Furthermore, using clever techniques for the compression will also reduce the average and peak power consumption.

The low power test data compression techniques can be divided into three main categories [22]: coding-based techniques, linear-decompression-based techniques (LFSR reseeding techniques) and broadcast-scan-based technique.

For example found in [23]. In the test cubes generated by ATPG are encoded using Golomb codes which are devel-
oped from run-length codes. All don’t care (X) bits are mapped to 0 and the Golomb code is used to encode runs of 0s. This technique is efficient to compress the runs of 0s but inefficient to encode the runs of 1s and the test storage may increase if there are many runs of 1s in the test cubes. The technique presented in [24] based on alternating run-length coding overcomes this problem.

In [25] a method that uses dual LFSR reseeding technique is used to reduce total power by masking some bits in the test vector by using AND and OR gates, thus reducing the transition probability. However, the method needs a hardware area overhead and increases test storage since there are two LFSRs used in the technique. In [26] the authors present a scheme that uses a LFSR reseeding technique to reduce the number of transitions in the scan-chain with an acceptable area overhead and good test power reduction.

5.12. Bit-Swapping Algorithm for Low-Power BIST

A modified linear feedback shift register (LFSR) is presented that reduces the number of transitions at the inputs of the circuit-under-test by 25% using a bit-swapping technique. They also show that the proposed design can be combined with other techniques to achieve a very substantial power reduction of up to 63%.

An important application of Linear Feedback Shift Registers (LFSRs) is in the generation of test vectors for digital circuits. This is because, with little overhead in the hardware area, a normal register can be configured to work as a test generator and, with an appropriate choice of the tap sequence (XOR locations), the LFSR can generate all possible output test vectors. Furthermore, the pseudorandom behavior of the LFSR reduces the correlation between successive test vectors, which means that it can achieve a high fault coverage in a relatively short run of test vectors. However, this lack of correlation substantially increases the Weighted Switching Activity (WSA) within the Circuit-Under-Test (CUT). This often causes the power consumed during test mode operation to be much higher than during normal mode operation, which can lead to problems with battery lifetime and system reliability.

It is therefore desirable to find a modification to the LFSR structure that can re-order the outputs in a manner that reduces switching activity without compromising the fault coverage. A modified LFSR that reduces the number of transitions in the CUT inputs by 25% using a bit-swapping technique. The proposed design can be achieved using a conventional LFSR and extra 2 x 1 multiplexers as shown in figure 4. The main advantage of the proposed design is the flexibility to be combined with other low-power techniques for further savings in power consumption.

5.13. Adaptive Shift Power Control technique

To reduce the scan shift power consumption in logic BIST by using highly correlated test stimulus bits among adjacent scan cells, all existing methods only manipulate test stimulus sequences generated by LFSR in various ways and the test responses are ignored completely. Although it has been observed that the Hamming distance between a test stimulus and its captured test response is typically small, the test stimulus of a test pattern is loaded into the scan chains at the same time as the test response of the previous test pattern is unloaded from the scan chains. Therefore, they are irrelevant. As a result, the scenarios listed below may happen when shifting the test stimulus and the test response simultaneously:

- The number of transitions contributed by the test response is much larger than that contributed by the test stimulus. [27]

5.14. Increasing Encoding Efficiency of LFSR Reseeding-Based Test Compression

Usually, the deterministic test set to be encoded by LFSR reseeding tends to have a biased probability for the logic value 1 or 0 at each primary input. The biased inputs are fixed to the logic value 1 or 0 with some combinational logic, so that the amount of data to be encoded by the LFSR can considerably be reduced. The combinational logic for bit fixing has to set some primary input to the logic value 0 (or 1), if the corresponding probability of the
6. Conclusion

In this paper a qualitative survey on low power testing techniques and its methodology was carried out. While analyzing, all dimensions of power during chip testing was considered as parameters. Further this paper gives a survey not only on algorithmic side but also on hardware approaches.

References


