A New Scan Architecture for Both Low Power Testing and Test Volume Compression Under SOC Test Environment

Hong-Sik Kim · Sungho Kang · Michael S. Hsiao

Received: 10 April 2007 / Accepted: 2 January 2008 / Published online: 18 January 2008 © Springer Science + Business Media, LLC 2008

Abstract A new scan architecture for both low power testing and test volume compression is proposed. For low power test requirements, only a subset of scan cells is loaded with test stimulus and captured with test responses by freezing the remaining scan cells according to the distribution of unspecified bits in the test cubes. In order to optimize the proposed process, a novel graph-based heuristic is proposed to partition the scan chains into several segments. For test volume reduction, a new LFSR reseeding based test compression scheme is proposed by reducing the maximum number of specified bits in the test cube set, s_{max} , virtually. The performance of a conventional LFSR reseeding scheme highly depends on s_{max} . In this paper, by using different clock phases between an LFSR and scan chains, and grouping the scan cells by a graphbased grouping heuristic, s_{max} could be virtually reduced. In addition, the reduced scan rippling in the proposed test compression scheme can contribute to reduce the test power consumption, while the reuse of some test results as the subsequent test stimulus in the low power testing scheme can reduce the test volume size. Experimental results on the largest ISCAS89 benchmark circuits show that the proposed technique can significantly reduce both the average

Responsible Editor: S. Hellebrand

H.-S. Kim (⊠) · S. Kang
Department of Electrical and Electronic Engineering,
Yonsei University,
Shinchondong, Seodaemungu,
Seoul, Korea
e-mail: hongsik@yonsei.ac.kr

M. S. Hsiao Bradley Department of Electrical and Computer Engineering, Virginia Tech., Blacksburg, VA, USA switching activity and the peak switching activity, and can aggressively reduce the volume of the test data, with little area overhead, compared to the previous methods.

Keyword System on a chip · Scan testing · Low power testing · Test compression

1 Introduction

As SOC (system on a chip) design trend continues to take root in the semiconductor industry, the complexity and the cost of testing the manufactured chips have also increased. The scan-based design style is an attractive solution to achieve desired fault coverage and to speed up time-tomarket in spite of its area and performance overhead [1, 12]. However, in scan testing, the shift operations to load and observe test data can lead to excessive transitions in the scan chain. In addition, unnecessary switching activity in the circuit-under-test (CUT) [2] and the reduced correlation between consecutive test pattern and test response can aggravate the power consumption during the capture cycle. The increased power dissipation may reduce the reliability of circuits and can increase the package cost because the heat dissipation and current density due to high switching activity in the circuit can exceed the limit of the design specification. Therefore, many techniques to reduce the test power in scan-based test environment have been proposed [5, 7, 13, 22–24, 28, 30, 31]. In [5, 22, 30], low power scan test techniques to reduce the transitions in scan-in vectors are studied. However, in these techniques, the transitions in scan-out vectors cannot be controlled and the peak power consumption during the capture cycle is not guaranteed. In [7, 24], the transitions in the scan chain have been prevented from propagating to circuit lines by adding externally controlled gates between the outputs of scan cells and the inputs of circuit under test. Although this technique can disable the switching activity inside the circuit-undertest completely during the scan shift operation, it can introduce undesirable timing impact on critical paths, and the power consumption in the capture cycle is not considered. In [23, 28, 31], a scan chain is partitioned into multiple segments, and only one segment is activated at a time. In [31], the original scan chain partition scheme has been proposed and it has been shown theoretically that the reduction of test power dissipation is linearly proportional to the number of segments without increasing the test application time. In [23], selective activation scheme is applied not only to the shift operation but also to the capture operation in order to reduce the peak power consumption by using different clock phases for segments in capture cycle. In this scheme, several clock cycles are required for exclusively activating segments in the capture cycle that results in higher testing application time. In [28], the test response data captured in some scan segments are used for the generation of the subsequent test stimulus by exploiting *don't care* bits in order to reduce both test time and test data volume. In [13], the test set is generated and ordered in such a way that some of the scan chains can be frozen for portions of the test set.

In addition to test power problem, the SOC design requires very large volume of test data because it integrates so many IP cores inside. In order to apply the large volume of test patterns to a chip under test, the automatic test equipment (ATE) requires a large size of memory storage so that the cost of the tester increases. Additionally, the limitation on the number of external IO ports of a chip and ATE makes its testing a very time consuming task. Therefore, in order to reduce the high test cost caused by the large test data volume, many techniques on compressing deterministic test set have been proposed. One class of test compression schemes is to encode the original test cubes by using a variety of codes [4, 6, 10, 11, 19, 29]. Theses schemes include run-length codes [6, 10], variablelength codes [11], fixed-length codes [29], Golomb codes [4], and code-based compression using evolutionary algorithm [19]. The next class is to use LFSR reseeding [9, 14, 15, 16, 21]. The original test compression methodology using LFSR reseeding was proposed in [9]. A seed is loaded into an LFSR and then the LFSR is run in an autonomous mode to fill a set of scan chains with a deterministic test pattern. If the maximum length of each scan chain is L, then the LFSR is run for L cycles to fill the scan chains. If a set of deterministic test cubes (test patterns where bits unassigned by ATPG are left as don't cares) is given, a seed can be computed for each test cube by solving a system of linear equations based on the feedback polynomial of the LFSR. In [9], it is estimated that in order to reduce the probability of not finding a solution for the system of linear equations to less than 10^{-6} , the LFSR degree should have a length of $s_{max}+20$, where s_{max} denotes the largest number of specified bits in any test cube in the test set. In order to improve the encoding efficiency of the basic LFSR reseeding scheme by reducing the dependency on s_{max} , many methods have been proposed [14, 15, 16, 21]. In [17], a test cube is divided into blocks by an encoding algorithm and the blocks that contain transitions are produced by LFSR reseeding. For the blocks that do not contain transitions are fed with a constant logic value. Since this scheme can decrease both the specified bits and the transitions, the test volume and the test power consumption can be reduced at the same time. However the peak power dissipation during a capture cycle is not considered.

This paper proposes a new scan based test methodology to reduce both the test power consumption and test data volume. For the test power optimization, a new scan architecture to reduce the power dissipation during both the shift and the capture cycles without impacting on the original fault coverage is proposed. In the proposed architecture, a scan chain is partitioned into several segments and the scan chain rippling is restricted to only one of the segments during scan shift operation at any given time. In addition, according to the profile of don't-care bit positions in the test cube set and the test response set, some segments are disabled and skipped from loading test data in order to reuse the test response data as its subsequent test stimulus so that the average power consumption can be further reduced. That is, if the captured data in a segment is the same as the consecutive test stimuli and do not contribute to the detection of the target faults, the segment does not load test data during scan shift period. In addition, some segments are disabled during capture period in order to reduce the peak power consumption during capture operation. The original test cube set is preserved in the final test stimulus data applied by the proposed scheme so that the fault coverage does not change. A novel graph-based heuristic for scan chain partitioning is devised in order to partition the scan chain into multiple segments efficiently so that the power dissipation during scan testing can be minimized.

For test data reduction, a new test compression scheme using LFSR reseeding is proposed. Since the efficiency of the traditional LFSR reseeding depends on s_{max} , the proposed scheme reduces s_{max} virtually by using different clock phases between an LFSR and scan chains, and grouping scan cells in each segment. If the clock frequency for the LFSR is slower than the clock frequency for the scan chain by r times, then the successive r scan cells are filled with the same data so that the number of specified bits can be reduced with an adequate grouping of scan cells. The real number of specified bits in the test cube set does not change, but the informative data for the seed calculation will be decreased so that s_{max} can be reduced *virtually*. A graph-based heuristic is used in order to group the scan cells.

In addition, the reduced scan rippling in the proposed test compression scheme reduces the average test power consumption, while the reuse of some test results as the subsequent test stimulus in the test power reduction scheme also reduces the test volume size. Experimental results on the largest ISCAS 89 benchmark circuits show that the proposed test power reduction scheme can improve the efficiency of the compression scheme by 13.65% on average. Meanwhile, the proposed compression scheme provides additional reduction of average power consumption by an average of 23.70%. When combining test power reduction and test compression, the complete architecture can reduce the average power consumption on average by 92.90% and the peak power consumption by 46.40%. Finally, the proposed architecture requires much less test storage compared to the previous test compression scheme with little hardware overhead.

The remainder of the paper is organized as follows. Section 2 describes the proposed scan architecture for the low power testing, including the scan chain partitioning heuristic and the method to increase the number of *don't care* bits in the test cube set. In Section 3, the proposed test compression scheme will be described. Experimental results will be provided in Section 4. Finally Section 5 concludes this paper.

2 Proposed Scan Architecture for Low Power Test

Figure 1 shows the proposed scan segmentation architecture and the detailed scheme for the scan segment and control unit. The proposed architecture is a little bit similar to the Illinois Scan Architecture (ILS) [8, 18]. ILS consists of two modes of operation; the one is the broadcast mode and the other is the serial mode. In the serial mode, it operates as a traditional scan chain, while in the broadcast mode, the scan chain is split into several length-balanced segments and the same data are broadcasted into each segment. That is, all the segments are activated and fed with the same data during shift cycle in broadcast mode. However in the proposed architecture, the scan chain is split into a given number of length-balanced segments and only one segment is activated or observed at a time during each test clock, in order to reduce the number of scan cells that are switching simultaneously during both scan shift and capture periods. In the shift cycle, the test data is scanned into one of the scan segments through the scan-in pin and the content of this segment is scanned-out through the scanout pin at the same time. The contents of the other partitions are frozen until the activated segment complete the entire shift operation. Prior to the activation of any particular segment, the segment configuration data is shifted into the scan configure register (SCR) through scan-in pin. Then, according to the contents of the SCR and the state of Test State Machine (TSM) which is a simple finite state machine (FSM), Segment Control Unit controls the activation of scan segments and the select lines of the multiplexer (MUX). In order to freeze the scan segments selectively, the scan cells are modified as shown in Fig. 1. To scan-in a new test data into the *i*th scan segment or scanout the test response from the *i*th scan segment, ScanEn shall be activated and Segment Disable_i shall be disabled. To capture the response to the scan cells in the *i*th scan segment, both ScanEn and Segment_Disable_i shall be disabled. If the signal Segment Disable_i is activated, the



Fig. 1 Proposed scan architecture

scan cells in the *i*th scan segment shall reserve their current contents or be frozen. In order to configure the SCR, additional 2N clock cycles are required (N for the shift disable setting and N for capture disable setting), where N is the number of scan segments. However, the total test time usually does not increase because the time reduction due to the segment disabling is much larger than the additional 2N clock cycles for SCR configuration.

Figure 2a illustrates the implementation of an example for a Test State Machine where the number of segments is three. It consists of some combinational logic (an offset decoder and a decoder) and a mod-N bit counter. In this example, 2-bit counter is required since the number of segments is three. During the shift period, the *i*th segment is skipped from activation if the *i*th SCR for shift disable is set to be 1. So the counter increases its current content by the amount of the offset which is pre-calculated with offset decoder. Figure 2b shows the simulation waveform for the example. During the test mode, three signals of *Segment* $Disable_i$ (i=1, 2, 3) for segment disable are generated for both shift and capture operation. At the first time, the scan chain is in segment configuration mode, and the segment configuration data are shifted into the SCR. Then the scan chain falls into scan shift mode and the scan segments are activated in the sequence of segments 1, 2 and 3 until the current test patterns are completely loaded into the scan chain. Since for the current test pattern, the second segments do not need to be loaded with test stimulus data (that is, the content of the second shift disable register in SCR is set), it is skipped from being activated and instead the third segment is controlled to be loaded with test data from the tester. For the second segment, it re-uses its current content as the consecutive test stimulus. After the scan chain is fully loaded, the signal of ScanEn is deactivated and it becomes the *capture mode*. In this example, the first segment does not need to capture the

test response data (that is, the content of the first capture disable register in SCR is set), so that it is disabled while the other segments are being controlled to capture the test response.

There have been several low power schemes based on scan segmentation method [23, 31]. In [23, 31], every segment loads serial data one by one during the scan shift cycle. In [23], in order to reduce the power consumption during capture cycle, only one segment captures the test response at a time. The difference between the proposed scheme and the previous ones is as follows. In the previous schemes, every segment is loaded with test data during scan shift period and captures test response data during capture period. However, in the proposed scheme, if the test response data captured in a segment is the same as the consecutive test stimuli and do not contribute to the detection of the target faults, they are reused as the next test stimulus data without being loaded with new test data. In addition, in case that the test response data to be captured in some segments are not of help for fault detection, the segments are disabled from capturing test response in order to reduce the power consumption during capture period. Therefore the maximum number of flip-flops which can toggle simultaneously is limited compared to the previous schemes so that both the average and the peak power dissipation during scan testing can be considerably reduced.

The proposed scan scheme makes use of the information about the distribution of *don't cares* in the given test cube set when partitioning the scan chain into multiple scan segments, reordering the test vectors, and controlling (disabling or activating) each scan segments. It fills the don't care bit positions in the consecutive test stimulus data with the test response data so that the original test cube set is preserved in the final test stimulus data. Therefore the proposed scheme does not degrade the original fault coverage.



unit

Depending on how the transition is launched and captured, there are two types of transition fault test pattern generation methods such as skewed load [25] and broadside method [26]. In the proposed scheme, the sequence of the activation of each segment is fixed. That is, only the scan cells in the last activated segment can launch the transition at the target gate in the last shift operation. Therefore the skewed load method cannot be applied in the proposed scheme. However, in case of broadside method, a pair of clock pluses is applied in order to launch and capture the transition at the target gate. Therefore, the broadside method can be applied in the proposed scheme. Obviously, a slight modification in the traditional transition fault ATPG process is required in order to apply the broadside method for the test of transition faults, since the proposed scheme exploits several techniques during ATPG such as freezing scan segments and reusing the test responses for the next test stimuli. However, how the ATPG process should be modified is beyond the scope of this particular paper.

2.1 The Scan Partition Heuristic

How to distribute scan cells to scan segments determines the efficiency of the proposed scheme. Without loss of generality, it would lead to better results to include the scan cells that will result in lower transition counts in the same segment and place the scan cells that will result in higher transition counts in the different segments in order to reduce the shift power consumption. In the proposed scan partition heuristic, the associated set of test cubes is analyzed by a graph-based heuristic. The nodes in the graph represent scan cells and the edge weights denote the number of test cubes where both of the corresponding two scan cells are specified and have the same test data. The edge weight thus determines the connection strength between the two nodes. It is desired to include highly connected nodes into the same scan segment, so that the average weight sum inside the segment should be kept high or the number of transitions inside the same scan segment can be kept low. Additionally the average weight sum among nodes across scan segments should be kept low, in order to place the scan cells that might result in higher transition counts in the different scan segments. Therefore, in this paper, a new heuristic for scan chain partition has been proposed in order to divide a graph into N sub-graphs with maximum weight sum inside the scan segment and with minimum weight sum crossing scan segments.

Figure 3 describes the heuristic where N is the number of scan segments and M is the balanced number of scan cells in each segments. First, a graph G is constructed from the given test cube set. A pair of nodes is selected among G such that its edge weight is maximal and is included into a sub-graph G_i . The selected pair is removed from G. Next, the node

```
construct a graph G;
for (i=1;i<N+1;i++)
{
    select the pair with the maximum edge weight sum from
    G;
    add the pair to group G;
    for (j=1;j<M+1;j++) // generating G;
    {
        select a node from G so that the weight sum of the
        edges between the node and G<sub>i</sub> is maximum;
        add the node to G<sub>i</sub>;
        remove the node from G;
        optimize_1 G;
    }
    optimize_2 G<sub>i</sub>;
}
```

Fig. 3 Scan partition algorithm

whose edge weight sum with G is maximal is selected and included into G_i , and then G_i is optimized. In the proposed heuristic, two optimization steps are applied in order to not only minimize the weight sum between sub-graphs but also maximize the weight sum inside each sub-graph. During the generation of G_i , the optimization process (optimize 1) has been applied so as to reduce the inter sub-graph weight sum. Therefore a selected node is replaced with an unselected node if the average edge weight sum between the current sub-graph and remaining sub-graphs is lower than the average edge weight sum between a new sub-graph including the unselected node and the other sub-graphs. After generating the sub-graph, the second optimization process (optimize 2) is applied to maximize the edge weight sum inside the sub-graph. Therefore an unselected node is replaced with an already selected node if the edge weight sum of the unselected node is higher than the weight sum of the selected node. This optimization is repeated until no more improvement is possible.

Figure 4 shows an example of the scan chain partition process. A graph *G* based on the given test cube set is constructed. In this example, the number of segments is assumed to be 2. Node 0 and node 3 yield the highest weight sum of 3 in *G* so that they are added to the sub graph G_1 . Since node 1 has the strongest connection or the maximum weight sum with the sub-graph G_1 , it is included into G_1 . Finally highly correlated scan cells are placed in the same segment. Nodes 0, 1, and 3 belong to one segment and the rest belong to the other segment.

Since some scan cells may not be included into the same segment because of layout constraints, the physical layout constrains should be considered in the scan partition process. Usually the layout is constructed so that the routing overhead shall be minimized. Therefore the nearby scan cells have to be placed in the same scan chain segment so as to avoid any layout violations in the implementation of an efficient scan chain partition.



2.2 Increasing the Number of *Don't Cares* in the Test Cube Set

The performance of the proposed scheme is sensitive to the initial set of test cubes. Since the proposed method utilizes the distribution of *don't care* bits in the test cube and the test response, if the number of *don't care* bits in the test cube set increases, then the scan partition becomes more efficient. Sometimes, however, most bits in the test pattern set may be specified. In such cases, in order to increase the number of *don't cares* in the test pattern set, some bit positions in the test patterns can be relaxed into don't-cares by a heuristic based on the concept of support sets [20].

A test vector is fully specified if all inputs are specified to 0 or 1 (i.e. no input assumes a value of X). A test vector that contains *don't cares* is said to be partially specified. A support set (*SS*) for a primary output Z is any set of signals including primary inputs (PIs) that satisfy all the following conditions.

- 1. All signals in the set assume a logic value 0 or 1.
- 2. The primary output Z is a member of the set.
- 3. The logic value on any signal except PIs in the SS is uniquely determined by values of other signals in the SS.

The support signals for a gate are the smallest subset of signals that are required to uniquely determine the current logic value of the gate. In case that multiple signals determine the logic value of the gate, the following criteria are used in order to calculate the smallest subset. If one of the possible support signals of the gate has already been included in the support set of the circuit, then this signal is selected as the support signal of the gate. Otherwise, a support signal at the

lowest level is chosen. For example, consider an AND gate a that has two inputs b and c. In case that signal b and c have the logic value 0 and 1 respectively, then the value on a is uniquely determined by the signal b so that the support signal for gate a is signal b. Assume that the logic values of signal b and c are all 0's, and the level of b is higher than the level of c. If b is already in the support set of the circuit, then b will be the support signal for a; otherwise, c will be the support signal of a, because the level of c is lower than b. In case of multiple primary outputs, condition 2 is modified to require that each of the POs be included in the support set. A support set is minimal if no signal in the set can be deleted without violating the conditions. It is desirable to compute a minimum support set, however it is very time consuming to compute the minimum support set for each input vector.

The procedure which takes a list of gates L with known logic values efficiently computes a minimum support set as shown in Fig. 5. By computing the support set for each test vector, the input sequence can be relaxed from fully specified sequence to partially specified sequence. In the process, the circuit is levelized and the input vector is simulated to determine the logic value of each circuit line. Then, the support sets are computed according to the logic values in circuit lines for each corresponding test vector.

With this technique to increase the number of *don't care* bits in the test data set, the dependency of the proposed low power scan architecture on the initial test data set can be reduced or the efficiency of the proposed scheme can be increased. We note that we perform test vector relaxation with the underlying fault coverage intact. In other words, the relaxed set of test cubes retains the original fault coverage.

3 A New Test Compression Scheme

Figure 6a shows the traditional LFSR based test compression scheme. A seed is loaded into an LFSR and then the

<pre>Procedure Compute_Support_Set(L)</pre>
{
support set $S = L$;
while (unsupported gates in L exists)
{
g = unsupported gate in L with maximal level;
ss = minimal support signals for g ;
add support signals ss to S;
for (all the unsupported gates <i>i</i> in <i>ss</i>)
{
add <i>i</i> to <i>L</i> ;
}
mark g supported;
}
return support set S;
}

Fig. 5 Procedure to compute support set

Fig. 6 Test compression schemes



b Proposed Reseeding Scheme

LFSR is run in an autonomous mode to fill a set of scan chains with a deterministic test pattern. If the maximum length of each scan chain is L, then the LFSR is run for Lcycles to fill the scan chains. Since different seeds generate different test patterns, if given a set of deterministic test cubes (test patterns where bits unassigned by ATPG are left as *don't cares*), the corresponding seeds can be computed for each test cube by solving a system of linear equations based on the feedback polynomial of the LFSR.

The proposed test compression scheme is shown in Fig. 6b. The proposed scheme is to reduce the maximum number of specified bits in the test cube sets virtually by using different clock phases between an LFSR and scan chains. In the traditional LFSR reseeding based test compression schemes, the LFSR and scan chain have the same clock frequency. Therefore all the scan cells have their own distinctive linear equations determined by the LFSR polynomial. However if we feed a slower clock to the LFSR by r times than the clock to the scan chain with adequate scan cell re-ordering (ϕ for scan chain and ϕ/r for the LFSR as shown in Fig. 1b, then the r consecutive scan cells shall have the same linear equations. By using different clock phases between an LFSR and scan chains, the length of the scan chain can be virtually reduced so that the number of specified bits can be decreased. The real number of specified bits in the test cube set does not change, but the data for the seed calculation reduce so that the s_{max} can be *virtually* reduced. Since the reduced number of specified bits means the reduced seed length, the encoding efficiency of LFSR reseeding scheme could be increased. Actually, since the proposed scheme feeds same test data into the consecutive scan cells, it is a little bit similar to the Illinois Scan Architecture (ILS) [8, 18] where the same test data is fed into the parallel scan chains. The difference of the proposed scheme is that the proposed scheme broadcasts the same data bit in time to consecutive scan cells belonging to the same party while ILS architecture broadcasts the same data bit in space to parallel scan chains. The efficiency of the proposed compression scheme depends on the order of scan cells and the pre-computed test cube set, so that if new test patterns are required after the design is manufactured, the new test patterns cannot be applied with the proposed scheme. In order to address this problem, the new test patterns should be directly applied to the scan chain from ATE without being compressed. In addition, since the proposed scheme requires partitioning of existing scan chains and reordering of scan cells according to an analysis on the test cubes and the test response data, the proposed scheme cannot be applied to the design where only hard cores are available.

3.1 The Scan Cell Grouping Heuristic

Since the whole analysis of the test cube set and scan cell grouping is not suitable for physical scan cell ordering. Therefore in the proposed scheme, only the scan cells in the same segment, which is divided by the scan partition heuristic for low power application as explained in Section 2.2, are grouped together into multiple cell groups where each scan cell has no conflict with each other. In order to identify the scan cells that belong to the same group, the associated set of test cubes is analyzed by a graph-based heuristic. The nodes in the graph represent scan cells in the same scan segment and the edge between two nodes means that there is no conflict of data contents between each other so that it means that two nodes connected with the edge can potentially be grouped together. The node strength is determined by the sum of the edges in each node. In order to divide a graph into N sub-graphs with the same number of scan cells, a new heuristic for scan chain partition has been used.

Figure 7 describes the heuristic where M is the balanced number of scan cells in each group. First, a graph G is constructed for the scan cells in a segment by analyzing the given test cube set. A pair of nodes is selected among Gsuch that the sum of their node strength is minimal and is included into a sub-graph G_i . The selected pair is removed from G. Next, the node whose node strength is minimal and which has connections with all the nodes in the sub-graph G_{i} is selected and included into G_{i} . This is repeated until there is no node which can be included into G_i or the number of nodes in G_i becomes M. After generation of G_i , the size of the sub-graph is balanced according to the size of the previously generated sub-graphs. Since the heuristic starts from nodes whose node connections are minimal and the size of the new sub-graph is maximally restricted to M, the size of the newly generated sub-graph, G_i , shall be smaller than or the same as the size of the previously generated sub-graph. In case that the size of G_i is the same as the size of the previous sub-graph, then no size balance is required. In case that the size of G_i is smaller than the size of the previous sub-graph, the nodes where all the test data contents are *don't cares* are included into G_i in order to adjust the size of G_i the same as the size of the previous

```
construct a graph G for scan cells in a segment;
G_{org} = G;
while (G \mathrel{!=} \phi)
  select the pair with the minimum sum of node connections from G;
  add the pair to group G<sub>i</sub>;
  remove the pair from group G:
  for (j=1;j<M+1;j++) // generating G_i
    select a node from G such that the node is connected to all the nodes in
    G_i, and its connection strength is minimal;
    add the node to Gi:
    remove the node from G:
  if(balance\_size(G_i) == false)
    decrease M;
                       // restart the process
    G = G_{org};
repeat the grouping process for the other segment
```

Fig. 7 Scan cell grouping algorithm

sub-graph. If there are no nodes available whose test data contents are all *don't cares*, the size of the sub-graph, M, is decreased and the process restarts. If the grouping is done successfully for a segment, this process is performed for the other segment. This is repeated for all the segments.

Figure 8 shows an example of the scan chain partition process. A graph G for the scan cells in a segment is constructed based on the given test cube set. Nodes 0 and 5 yield the smallest sum of connection strengths of 4 in G so that they are added to the sub graph G_1 . Since node 3 is the only node which has connection with all the nodes in the sub-graph G_1 , it is included into G_1 . Finally a new group is generated and its size is 3. This is repeated until every node in G is grouped together. Nodes 0, 3, and 5 belong to one group, and nodes 2, 4 and 7 belong to another group. The rest belong to the third group. In this example, the third group, or the sub-graph 3, has only two nodes, while others have three nodes. In this case, in order to avoid inserting dummy cells in that segment to balance the size of segments, the scan cells in that segment are placed closest to the scan in port as shown in Fig. 8.

4 Experimental Environment and Results

4.1 Experimental Environment

Figure 9 shows the total architecture for the proposed low power testing and test data compression. During the shift period, the encoded data is loaded into LFSR through a scan in pin. After the LFSR is completely loaded with the encoded data, the SCR configuration data is decoded from LFSR and fed into SCR register. SCR register generates two signals such as ShiftDS and CaptureDS for each segment in order to let TCU (test control unit) know whether the corresponding segment should be disabled or not. And then test data is decoded and fed into each segment from LFSR, while TCU generates segment enable signal for each segment one by one according to the signal ShiftDS and CaptureDS from SCR. The input data length for each test pattern is the same as the LFSR length. As explained earlier, the efficiency of the proposed compression scheme depends on the order of scan cells and the precomputed test cube set, so that if new test patterns are required after the design is manufactured, the new test patterns cannot be applied with the proposed scheme. In order to address this problem, a mux is added between LFSR and scan chain to apply the uncompressed new test patterns directly to the scan chain from ATE.

Experiments were conducted on the large ISCAS89 benchmark circuits [3]. Only the flip-flops in the benchmark circuits were considered as the scan cells, while the primary inputs and primary outputs were assumed to be



fully controllable and observable by the external tester and were not considered during partitioning the scan chain and grouping the scan cells in each segment. Since some scan cells may not be included into the same segment because of layout constraints, the physical layout constraints should be considered in the scan partition process. Usually the layout is constructed so that the routing overhead shall be minimized. Therefore the nearby scan cells have to be placed in the same partition so as to avoid any layout violations in the implementation of an efficient scan chain partition.

For each circuit, deterministic ATPG, which is based on the SOCRATES algorithm [27], was performed to generate the deterministic test cubes for the detectable stuck-at faults in the benchmark circuits targeting at 100% fault coverage. We note that other ATPG's may be used as well. Based on the deterministic test cubes, the proposed low power test scheme (including the heuristic to increase the number of *don't cares* in the test cube and the segment partition

Fig. 9 The total architecture for both low power testing and test data compression

algorithm) and the proposed test compression scheme (including scan cell grouping algorithm and seed calculation) are applied to the benchmark circuits. Table 1 presents the hardware area of the proposed architecture, including LFSR, SCR, TCU, Counter and output Mux. The hardware area is given in terms of the number of gate equivalents, assuming that a two-input NAND gate is 1 gate equivalent. The second column includes the number of gate equivalents of ISCAS benchmark circuits. The remaining columns include the number of gate equivalents of the subfunctional blocks in the proposed reseeding architecture. The proposed architecture is implemented by synthesizing with Design Compiler from Synopsys. The last column includes the total area and the overheads calculated by dividing the total area of the proposed scheme with the area of the corresponding benchmark circuit. The overhead for each benchmark circuit is less than 6.9% in all cases. The average area overhead is about 2.38%. The results are based on 3 partitions.



Table 1 Hardware area of the proposed scheme

Circuit		Gate equivalent									Total area (% overhead)	
		LFSR		SCR		TCU		Count	er	Output mux		
Name	Gate equivalent	FF's	Gates	FF's	Gates	FF's	Gates	FF's	Gates			
s5378	4,271	98	12	44	20	21	15	54	22	7	293 (6.86%)	
s9234	8,579	322	12	44	20	21	15	81	36	7	558 (6.50%)	
s13207	14,260	140	12	44	20	21	15	63	27	7	349 (2.45%)	
s15850	16,280	217	12	44	20	21	15	72	31	7	320 (2.70%)	
s38417	38,011	469	15	44	20	21	15	81	36	7	708 (1.86%)	
s38584	37,554	259	12	44	20	21	15	72	31	7	481 (1.28%)	
Average	19,825	251	13	44	20	21	15	71	31	7	471 (2.38%)	

4.2 Experimental Results on Test Power Consumption

Since in CMOS technology the major power dissipation comes from the switching of a CMOS gate from one stable state to another, the switching activity of charging and discharging the load capacitance in each component is usually the dominant factor in dynamic power dissipation. Thus the power consumption is usually evaluated as follows.

$$P_d = 0.5 \times VDD^2 \times f_p \times \sum E[T_i] \times C_i$$

where VDD is the power supply voltage, f_p is the clock frequency, $E[T_i]$ is the expected number of transitions per cycle in gate *i*, and C_i is the load capacitance of gate *i*. The total power consumption can be simply evaluated by using the weighted switching activity (WSA) [2, 30]. The weighted switching activity in each gate is given by the number of signal transitions times one plus the number of fanouts at the gate, as follows.

Table 2 shows the experimental results of the proposed scheme on large ISCAS 89 benchmark circuits in terms of the average WSA and the peak WSA. The average WSA and the peak WSA can explain the average power consumption and peak power consumption, respectively. In this experiment, the proposed compression scheme is not applied. According to the distribution of *don't care* bit positions in the

given test cube set, the scan chain is split into some multiple segments (two, three or four segments in our work) and the actual test stimulus data are calculated by the proposed scan partition heuristic. The method to increase the number of *don't-cares* in the test cube set can improve the efficiency of the scan chain partition heuristic during the scan partition process.

The second and the third columns in Table 2 illustrate the average WSA and the peak WSA, respectively, for each circuit under full scan environment. The results for 2 segments, 3 segments and 4 segments are shown in the remaining columns. In the case of 2 segments, the average WSA has been reduced to about 11.36% (558/4914) and the peak WSA has been reduced to 56.07% (3785/6751) compared to the case of the full scan environment. In the case of 3 segments, the average WSA has been reduced into 8.47% (416/4914) and the peak WSA into 49.44% (3338/ 6751). In the case of 4 segments, the average WSA has been reduced into 6.76% (332/4914) and the peak WSA into 45.90% (3099/6751). The peak power consumption is usually determined by the capture operation in the scan based test and in the capture cycle more than 1 segment can be activated. Since only one segment is activated in the shift operation of the proposed scheme, the reduction of average power consumption is much higher than the reduction of peak power consumption as shown in Table 2.

Table 2 Experimental results of the proposed scheme on average and peak power consumption

Name	Full scan		2 Partitions		3 Partitions		4 Partitions	
	Ave. WSA	Peak WSA	Ave. WSA	Peak WSA	Ave. WSA	Peak WSA	Ave. WSA	Peak WSA
s5378	1,053	1,728	193	928	107	744	76	660
s9234	2,020	3,294	543	2,117	392	2,065	183	1,464
s13207	2,136	4,350	249	1,253	154	916	99	798
s15850	3,962	5,301	252	2,590	105	2,410	105	2,415
s38417	11,378	12,369	1,045	4,629	887	3,248	678	2,784
s38584	8,934	13,464	1,067	11,195	852	10,642	850	10,475
Average	4,914	6,751	558	3,785	416	3,338	332	3099

Name	Full scan	2 Partitions		3 Partitions		4 Partitions	
		Without compression	With compression	Without compression	With compression	Without compression	With compression
s5378	1,053	193	141	107	71	76	61
s9234	2,020	543	392	392	279	183	139
s13207	2,136	249	201	154	126	99	82
s15850	3,962	252	184	105	88	105	72
s38417	11,378	1,045	791	887	673	678	537
s38584	8,934	1,067	824	852	652	850	760
Average	4,914	558	505	416	384	332	326

Table 3 Average power reduction in case of applying the proposed compression algorithm

As the number of segments increases, the area overhead and power reduction of the proposed scheme increase. However, the reduction of the power consumption slows down when the number of segments is larger than five segments, so the results under four segments are presented in Table 2.

Table 3 shows the average test power consumption in case that the proposed compression scheme is applied. Since the proposed compression scheme is to reduce the clock



Fig. 10 Comparison to previous scan chain partition schemes. a Comparison of average power consumption. b Comparison of previous scan chain partition schemes

frequency of the LFSR, the data rippling in the scan chain is considerably reduced so that the average power can be reduced. According to the experimental results as shown in Table 3, the proposed compression scheme could contribute the additional reduction in the average power consumption in all cases. Since the proposed compression scheme reduce only the data rippling during the scan shift operation, it does not contribute the peak power consumption.

In addition, the large amount of the switching activities during the scan shift period can increase the voltage drop so as to give rise to an unexpected setup time violation in the scan shift operation. Therefore higher clock frequency cannot be used for the scan shift operation in the traditional scan architecture. However, in the proposed scheme, the switching activities in the circuit lines during the scan shift period significantly reduce, so that higher clock frequency can be used for the scan shift operation. According to Table 3, the average WSA of the traditional scan scheme is larger than the average WSA of the proposed scheme with two, three, and four segments by 9.73, 12.80 and 15.07 times, respectively. Therefore, theoretically the proposed scheme can use higher clock frequency for the scan shift operation by 9.73~15.07 times than traditional scan scheme.

Figure 10 illustrates the comparison of the proposed scheme with the previous scan partition schemes [13, 17, 23, 28, 31]. WSA in each scheme has been normalized to

 Table 4 Compression results with or without using low test power method

Name	Tests	Original	s _{max}	Test storage			
		storage		Without low power method	With low power method		
s5378	201	43,014	29	6,432	5,220		
s9234	261	64,467	53	13,050	11,280		
s13207	248	173,600	47	12,152	10,192		
s15850	239	146,029	51	13,154	10,600		
s38417	375	624,000	97	35,625	30,255		
s38584	264	386,496	87	23,496	22,204		
average	205	205,372	61	14,844	12,817		

Name	[15]		[16]		[6]		Proposed scheme	
	Tests	Storage	Tests	Storage	Tests	Storage	Tests	Storage
s5378	196	6,180	NA	NA	100	8,502	201	5,220
s9234	205	12,112	NA	NA	111	10,608	261	11,280
s13207	266	11,285	255	11,320	235	18,518	248	10,192
s15850	269	12,438	142	11,584	97	15,900	239	10,600
s38417	376	34,767	105	30,560	87	57,118	375	30,255
s38584	296	29,397	192	27,428	114	53,774	264	22,204
average	268	17,697	174	20,223	105	23,489	205	12,817

Table 5 Comparison of test storages for different schemes

the WSA of the traditional full scan environment. The number of the scan partitions is three for all the schemes. The comparison on average power is shown in Fig. 9a. For all the benchmark circuits, the proposed scheme consumes the least average power. In case of s38584, the power consumption of the proposed scheme is slightly less than the power consumption of [23]. On the average, the average power consumption of the proposed scheme is much less than the previous scan chain partition schemes. As shown in Fig. 9a, the variance in the power consumptions for benchmark circuits is much smaller than the other schemes. Figure 9b compares the peak power consumption. In [13, 17, 28, 31], the data for the peak power consumption are not available since those methods do not consider the peak power reduction. For most circuits, the proposed scheme consumes least power. On average, the proposed scheme shows better peak power consumption than [23] by approximately 15%.

4.3 Experimental Results on Test Compression

The proposed compression method has been implemented in C, and applied to ISCAS 89 benchmark circuits with/ without the proposed low power method applied. Table 4 shows the experimental results for both cases with/without the low power application. The first column of the table gives the circuit names. The second column lists the number of test patterns. The third shows the volume of the original test data. In, the forth column, the sizes of the s_{max} are given. The last column gives the volume of the compressed test data for each case. Since several test cubes are self-generated from the previous test responses in case that the proposed low power method is applied, we don't need to encode those test cubes nor store their seeds. Therefore the volume of the compressed test data can be reduced much further by applying the proposed low power method. According to the table, for all the benchmark circuits, there are additional reductions in the test storage. On average, applying the low power method could reduce the storage size by about 14%.

Table 5 shows the comparison of test storages between the proposed method and three previous compression methods. For each compression scheme, their test size (tests) and volume of the compressed data (storage) are listed. Since [6] do not use test cubes but test patterns, its test size is much smaller than the others. In all the benchmark circuits except s9234, the proposed scheme requires the smallest storage size of [6] is slightly smaller than the storage size of the proposed scheme. On average, the proposed scheme requires much smaller storage compared to the previous schemes by 28~45%.

Table 6 shows the comparison of *compression ratio* between the proposed scheme and the five previous compression schemes. The *compression ratio* can be calculated by dividing the total amount of storage required to explicitly store the deterministic test patterns (the product of the length of a scan chain and the number of test cubes) by the amount of the encoded test data. The *compression ratio* is better for higher values of this number. In case of s9234, s13207 and s15850, [4], [19] and [17] shows slightly higher *compression ratios* by only 0.7 point, 0.09 point and 0.26 point, separately. However, in other cases, the proposed method provides much higher *compression ratio* than the previous methods. On average, the proposed

 Table 6
 Comparison of compression ratio for different schemes

Name	[11]	[29]	[4]	[19]	[17]	Proposed Scheme
s5378 s9234 s13207 s15850 s38417 s38584	62.00 70.90 83.50 78.90 53.60 NA	51.64 50.91 82.31 66.38 60.63 65.53	62.24 NA 94.21 88.74 89.68 82.11	81.20 83.20 89.60 86.30 NA 90.00	NA 79 94 93 95 93	87.86 82.50 94.13 92.74 95.15 94.26
average	69.80	62.90	83.40	86.10	90.8	91.11

scheme shows higher compression ratio by 0.31~28.21 point compared to the previous schemes.

5 Conclusion

In this paper, a new test methodology for both test power reduction and test volume compression has been proposed. For low power application, a new architecture using both scan chain partitioning and disabling is proposed in order to reduce both the average and peak power consumption. The scan chain is split into several length-balanced segments and only one segment is enabled in each test clock during both shift and capture cycle. In addition, the segments that do not need to be activated are skipped or disabled during shift and capture cycle. Therefore with the proposed scheme, both the average power and peak power consumption during test application can be significantly reduced. Since the efficiency of the proposed scheme highly depends on how to distribute scan cells into scan segments, a new graph-based heuristic for scan partition has been developed so as to not only increase the average weight sum inside each sub-graph and but also reduce the average weight sum among sub-graphs. Also, since the proposed scheme is sensitive to the initial test data set, a method based on support set has been proposed in order to increase the number of don't care bits in a test data set.

For test compression, a new LFSR reseeding scheme is proposed by reducing the maximum number of specified bits in the test cube set, s_{max} , virtually. If the clock frequency for an LFSR is slower than the clock frequency for scan chains by r times, then the r successive scan cells are fed with the same data. Therefore the number of specified bits can be virtually reduced by efficiently grouping scan cells. The real number of specified bits in the test cube set does not change, but the informative data for the seed calculation will be decreased so the s_{max} can be reduced virtually. A graph-based heuristic is used to group the scan cells in the same segment. In addition, the reduced scan rippling in the proposed test compression scheme can contribute to reduce the average test power consumption, while the reuse of some test results as the subsequent test stimulus in the test power scheme can reduce the test volume size. The experimental results based on the larger ISCAS89 benchmark circuits show that the proposed scheme can consume much smaller average and peak power during scan testing, and also provide much less test data storage than the previous methods, with little hardware overhead.

Acknowledgments The authors would like to thank the anonymous reviewers and Prof. S. Hellebrand for their helpful comments and suggestions to improve this paper.

References

- 1. Abramovic M, Friedman AD, Breuer MA (1993) Digital system testing and testable design. Wiley, New York
- Basturkmen NZ, Reddy SM, Pomeranz I (2002) A low power pseudo-random BIST technique. Proceedings of the IEEE International Conference on Computer Design, pp 16–18
- Brglez F, Bryan D, Kozminski K (1989) Combinational profile of sequential benchmark circuits. International Symposium on Circuits and Systems, pp 1929–1934
- Chandra A, Chakrabarty K (2000) Test data compression for system-on-a-chip using golomb codes. Proceedings of the VLSI Test Symposium, pp 113–120
- Chandra A, Chakrabarty K (2001) A unified approach to reduce SOC test data volume, scan power, and testing time. IEEE Trans Comput-Aided Des Integr Circuits Syst 20:355–368
- 6. Doi Y, Jajihara S, Wen X, Li L, Charkrabarty K (2005) Test compression for scan circuits using scan polarity adjustment and pinpoint test relaxation. Proceedings of the of Asia and South Pacific Design Automation Conference, pp 59–64
- Gerstendorfer S, Wunderlich HJ (1999) Minimized power consumption for scan-based BIST. Proceedings of the International Test Conference, pp 77–84
- Hamzaoglu I, Patel JH (1999) Reducing test application time for full scan embedded cores. In Digest Papers, 29th International Symposium on Fault Tolerant Computing, pp 260–267
- Hellebrand S, Rajski J, Tarnick S, Venkataraman S, Coutois B (1995) Built-in test for circuits with scan based on reseeding of multiple-polynomial linear feedback shift registers. IEEE Trans Comput 44(2):223–233
- Jas A, Touba NA (1998) Test vector decompression via cyclical scan chains and its application to testing core-based design. Proceedings of the International Test Conference, pp 458–464
- Jas A, Pouya B, Touba NA (1999) Scan vector compression/ decompression using statistical coding. Proceedings of the IEEE VLSI Test Symposium, pp 114–120
- Jha NK, Gupta S (2003) Testing of digital systems. Cambridge University Press, Cambridge
- Kajihara S, Ishida K, Miyase K (2002) Test vector modification for power reduction during scan testing. Proceedings of the VLSI Test Symposium, pp 160–165
- Kim H-S, Kim YJ, Kang S (2003) Test-decompression mechanism using a variable-length multiple-polynomial LFSR. IEEE Trans VLSI Syst 11(4):687–690
- Krishna CV, Touba NA (2002) Reducing test data volume using lfsr reseeding with seed compression. Proceedings of the International Test Conference, pp 321–330
- Krishna CV, Touba NA (2004) 3-Stage variable length continuous-flow scan vector decompression scheme. Proceedings of the IEEE VLSI Test Symposium, pp 79–86
- Lee J, Touba NA (2007) LFSR-reseeding scheme achieving lowpower dissipation during test. IEEE Trans Comput-Aided Des Integr Circuits Syst 26(2):396–401
- Pandey AR, Patel JH (2002) An incremental algorithm for test generation in Illinois scan architecture based designs. Proceedings of the Design, Automation and Test in Europe Conference, pp 369–375
- Polian I, Czutro A, Becker B (2005) Evolutionary optimization in code-based test compression. Proceedings of the Design, Automation and Test in Europe Conference and Exhibition, pp 1124– 1129
- Raghunathan A, Chakradhar ST (1995) Acceleration techniques for dynamic vector compaction. Proceedings of the IEEE/ACM International Conference on Computer-Aided Design, pp 310– 317

- Rajski J, Tyszer J, Zacharia N (1998) Test data decompression for multiple scan designs with boundary scan. IEEE Trans Comput 47 (11):1188–1200
- Rosinger PM, Gonciari PT, Al-Hashimi BM, Nicolici N (2001) Simultaneous reduction in volume of test data and power dissipation for systems-on-a-chip. Electron Lett 37(24):1434–1436
- Rosinger P, Al-Hashimi BM, Nicolici N (2004) Scan architecture with mutually exclusive scan segment activation for shift and capture power reduction. IEEE Trans Comput-Aided Des Integr Circuits Syst 23(7):1142–1153
- 24. Sankaralingam R, Touba NA (2002) Inserting test points to control peak power during scan testing. Proceedings of the IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems, pp 138–146
- Savir J (1992) Skewed-load transition test: part i, calculus. Proceedings of the International Test Conference, pp 705–713
- Savir J, Patil S (1994) On broad-side delay test. Proceedings of the VLSI Test Symposium, pp 284–290
- 27. Schulz M, Trischler A, Sarfert T (1987) Socrates: a highly efficient automatic test pattern generation system. Proceedings of the International Test Conference, pp 1016–1026
- Sinanoglu O, Orailoglu A (2002) A novel scan architecture for power-efficient, rapid test. Proceedings of the IEEE/ACM International Conference on Computer Aided Design, pp 299– 303
- Tehranipoor M, Nourani M, Chkrabarty K (2005) Nine-coded compression technique for testing embedded cores in SoCs. IEEE Trans Very Large Scale Integr (VLSI) Syst 13(6):719–731
- Wang S, Gupta SK (1997) DS-LFSR: a New BIST TPG for low heat dissipation. Proceedings of the International Test Conference, pp 848–857
- Whetsel L (2000) Adapting scan architecture for low power operation. Proceedings of the International Test Conference, pp 863–872

Hong-Sik Kim received the B.S., M.S., and Ph.D. degrees in Electrical and Electronic Engineering from Yonsei University, in 1997, 1999, 2004, respectively. He was a Post-Doctorial Fellow at Virginia Tech, in 2005, VA, and a Senior Engineer at System LSI

Group in Samsung Electronics Co, in 2006. He is currently a research professor at Yonsei University, Seoul, Korea. His current research interest includes design for testability, built in self test, and test compression algorithm.

Sungho Kang received the B.S. degree from Seoul National University, Seoul, Korea, and M.S. and Ph.D. degree in electrical and computer engineering from University of Texas at Austin. He was a Post-Doctorial Fellow at University of Texas at Austin, a Research Scientist at the Schlumberger Laboratory for Computer Science, Schlumberger Inc., and a Senior Staff Engineer at Semiconductor Systems Design Technology, Motorola Ins., Since 1994, he has been an Associate Professor at Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea. His current research interests include VLSI design, VLSI CAD and VLSI testing and design for testability.

Michael S. Hsiao received the B.S. degree in computer engineering (with highest honors) from the University of Illinois, Urbana-Champaign, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from the same university, in 1993 and 1997, respectively. In the summer of 1997, he was a Visiting Scientist at NEC, Princeton, NJ. Prior to joining Virginia Polytechnic Institute and State University, Blacksburg, (Virginia Tech) in 2001, he was an Assistant Professor with the Department of Electrical and Computer Engineering at Rutgers University. During the summer of 2002, he was a Visiting Professor at Intel in Santa Clara, CA. He is currently an Associate Professor in the Bradley Department of Electrical and Computer Engineering, Virginia Tech. He has published more than 120 refereed journal and conference papers. His current research interests include on test and verification of both nanoscale and VLSI systems, as well as diagnosis and power management of these systems. During his studies, Dr. Hsiao was a recipient of the Digital Equipment Corporation Fellowship, McDonnell Douglas Scholarship, and Semiconductor Research Corporation Research Assistantship. He is a recipient of the National Science Foundation CAREER Award, and he is a recipient of the Dean's Faculty Fellow at Virginia Tech. He serves on the editorial board for two journals and on program committee for several conferences and workshops.