

POWER AWARE TEST-DATA COMPRESSION FOR SCAN-BASED TESTING

G. Gekas¹, D. Nikolos¹, E. Kalligeros¹ and X. Kavousianos²

¹Computer Engineering & Informatics Dept., University of Patras, 26500, Greece

²Computer Science Dept., University of Ioannina, 45110, Greece

gekas@ceid.upatras.gr, nikolosd@cti.gr, kalliger@ceid.upatras.gr, kabousia@cs.uoi.gr

ABSTRACT

In this paper a new approach that targets the reduction of both the test-data volume and the scan-power dissipation during testing of a digital system's cores is proposed. For achieving the two aforementioned goals, a novel algorithm that inserts some inverters in the scan chain(s) of the core under test (CUT) is presented. However, no performance or area penalty is imposed on the CUT since, instead of additional inverters, the negated outputs of the scan flip-flops can be utilized. The proposed algorithm targets the maximization of run-lengths of zeros (or ones) in the test set accompanying the CUT. This algorithm combined with the Minimum Transition Count mapping of don't cares in a test set as well as with the alternating run-length code that have been recently proposed, achieves better test-data compression and reduced scan-power results than the relative works in the literature.

1. INTRODUCTION

The ever-increasing size, density and clock frequencies of contemporary Systems-on-a-Chip (SoCs) pose several difficult test challenges. The prevalent, core-oriented design style, although reducing the time-to-market and the complexity of the designers' task, leads to circuits with reduced accessibility and increased test-data-storage, test-application-time and test-power requirements. Consequently, the introduction of new, test resource partitioning (TRP) solutions that overcome these problems is of great importance.

The vast majority of the techniques that have been presented in the literature so far, focus solely on the problem of increased test-data volume or on that of increased test-power dissipation, ignoring the other. Various techniques have been thus proposed for test-data compression. Most of them are based on the use of different codes such as Golomb [1], Frequency Directed Run-length (FDR) [2], selective Huffman [3], or on linear test-cube decompression (performed by either a Linear Feedback Shift Register -LFSR- [4] or other combinational structures [5]). As for the minimization of testing power, several methods have been also proposed [6]-[11]. The problem of long test-application times is usually tackled by reducing the size of the test sets that should be applied to the various cores (e.g., by applying dynamic compaction during automatic test pattern generation -ATPG- [12]).

However, there are only few methods that handle both issues of increased data-volumes and power-dissipation during testing [13], [14]. The reason is that, most of the times, these two targets have contradictory requirements. For example, the reduced transition-count prerequisite for lowering scan-in power dissipation, makes classical run-length or LFSR-based compression inefficient. The most recent and successful technique, which provides a unified solution, is that of [14]. There, the Minimum Transition Count (MTC) mapping of a test set's don't cares [15] is combined with a new, modified FDR code called alternating run-length. Thus, reduced scan-power and test-data-volume results are achieved.

In this work, a novel approach that tackles both the aforementioned problems is presented. The proposed technique is based on the proper insertion of some inverters in the scan chain(s) of the CUT, in order to modify its original test set in such a way that will allow for better compression and reduced scan-power dissipation. Scan power is the dominant contributor to power dissipation during testing, as reported in [9]. We should note that the proposed method does not impose any hardware or performance overhead on the CUT, since, instead of inserting additional inverters in the scan chain(s), the negated scan flip-flop outputs can be utilized. The idea of inverting parts of the scan-chain for reducing testing power has been previously presented in [11]. In that work, the transition frequency between two test-set columns was used for determining the inverter-insertion positions, while scan chain rearrangements were performed for further reducing power dissipation. However, scan-chain reorganization may not be acceptable due to area, performance or production-cost reasons. This is why the proposed method leaves the scan chains of the CUT intact and introduces a new inverter-insertion criterion that, except for power minimization, targets also efficient test-data compression, which is not the case for the technique of [11]. The whole algorithmic framework of the proposed approach is presented in the following section, while experimental results and comparisons are provided in Section 3. The paper is concluded in Section 4.

2. SCAN-CHAIN INVERTER-INSERTION ALGORITHM

The proposed scan-chain inverter-insertion algorithm targets the minimization of the test data that will be stored in the tester, as well as the average power that will be dissipated during the scan-in operation. It receives as input a partially specified test set T_D , it calculates the posi-

We thank the European Social Fund (ESF), Operational Program for Educational and Vocational Training II (EPEAEK II), and particularly the Program PYTHAGORAS, for funding the above work.

tions of the scan chain(s) of the CUT in which inverters should be placed and outputs a compressed set of data T_E . T_E is stored in an external tester and is fed to an on-chip decoder. After decoding and shifting T_E in the scan chain(s) of the CUT, the original test set emerges (with its don't care bits filled with 0 or 1, of course). We should note that the proposed technique does not change the order of the test cubes (test vectors with 'x' values) of T_D and as a result, it can be applied to test sets of either full- or non full-scan cores.

A wide variety of codes have been used in the literature for compressing the test set of a CUT. From these codes, some run-length variants (Golomb, FDR) have been proven very efficient, while, at the same time, impose negligible area overhead on the CUT (i.e., the decompression circuit is very small and independent of T_D). In addition, the use of a code that belongs to the family of run-length codes is well suited for the case of low-power testing, since the optimization of a test set so as to be best compressed by such a code, results in long runs of zeros (or ones), which in turn lead to a reduction of the power dissipated during testing. For that reason, the use of a run-length-code variant is a very good choice for the problem at hand. In the following we will present our method assuming that the CUT has a single scan chain. Its application to the multiple-scan-chains case is straightforward.

The main idea of the proposed algorithm is to insert inverters before the inputs of some cells of the scan chain of the CUT so as to modify T_D in such a way that the resulting (modified) test set T_D' has longer runs of 0s (or 1s) than the original test set (T_D). Thus T_D' can be compressed more efficiently than T_D by a run-length-type code. We will next describe our technique targeting the maximization of run-lengths of 0s. However, a similar approach can be followed for maximizing run-lengths of 1s.

A simple yet very effective criterion is introduced for deciding if scan cell i should receive an inverted input or not. We check the corresponding (i th) column of T_D and if the percentage of 1s among the defined bits of that column is greater than 50, then an inverter should be placed in the scan chain before cell i (or equivalently, cell i should be driven by the inverted output of cell $i-1$). Of course, another inverter should be placed after cell i for providing the next scan cells with uninverted bits. If the values of a group of adjacent columns have to be inverted, then just two inverters need to be placed in the scan chain, one before and one after the group of the corresponding (adjacent) cells. The example of Figure 1 will clarify the proposed scan-chain inverter-insertion algorithm.

| | First bit entering the scan chain | Example test set T_D | | | | | | |
|-------|-----------------------------------|------------------------|-----|----|---|------|----|----|
| v_1 | 0 | x | x | 1 | 0 | 1 | 1 | 0 |
| v_2 | 0 | x | x | 1 | 0 | x | 0 | 1 |
| v_3 | 0 | x | 1 | 1 | x | x | x | 0 |
| v_4 | 1 | x | x | 0 | 0 | 1 | x | 0 |
| v_5 | 0 | 1 | x | x | x | x | x | 1 |
| v_6 | 1 | x | x | 0 | x | 0 | x | 1 |
| 1's % | 33.3 | 100 | 100 | 60 | 0 | 66.7 | 50 | 50 |

Figure 1. Calculating the 1s percentage for each column

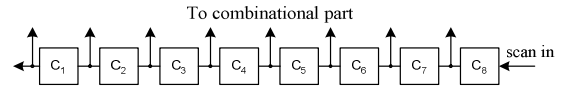


Figure 2. Example scan chain

In the above example we consider a test set of six test cubes (v_1 to v_6) and for each one of the eight columns of T_D we calculate the percentage of 1s among the defined bits of the corresponding column (the scan chain of the CUT before the inverter insertion is shown in Figure 2). As we can see in Figure 1, for columns 2, 3, 4 and 6 this percentage is greater than 50. Therefore, according to the proposed criterion, an inverter should be placed before scan cell 4 and after scan cell 2 (for the inversion of the group of columns 2 - 4), while two more inverters are required before and after scan cell 6 (for the inversion of column 6). Thus, test set T_D is modified as shown in Figure 3. The corresponding scan-chain modifications are presented in Figure 4.

| | Modified test set T_D' | | | | | | | |
|-------|--------------------------|---|---|---|---|---|---|---|
| v_1 | 0 | x | x | 0 | 0 | 0 | 1 | 0 |
| v_2 | 0 | x | x | 0 | 0 | x | 0 | 1 |
| v_3 | 0 | x | 0 | 0 | x | x | x | 0 |
| v_4 | 1 | x | x | 1 | 0 | 0 | x | 0 |
| v_5 | 0 | 0 | x | x | x | x | x | 1 |
| v_6 | 1 | x | x | 1 | x | 1 | x | 1 |

Figure 3. The resulting modified test set T_D'

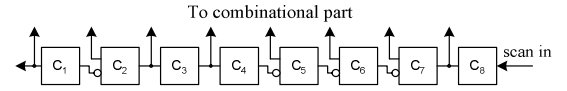


Figure 4. The modified scan chain

The average run-length of 0s for the original test set of Figure 1 (assuming that the 'x' bits are set to 0) is equal to 2.69, while for the modified test set of Figure 3 the same average run-length reaches 4.33. It is obvious that the proposed criterion increases the run-lengths of 0s in T_D' , leading this way to better compression and smaller power dissipation results.

It remains to explain how we set the values of the 'x' bits of T_D' . Zero filling is the obvious approach. That is, since we have chosen to create as long run-lengths of 0s as possible by inverting some portions of the scan chain of the CUT, then we could set the 'x' bits to zero value so as to further favorite this choice. However this is not the best solution from the perspective of power, since there will probably be several 'x' bits that will be preceded and/or followed by 1s. A more efficient approach would be to apply the MTC (Minimum Transition Count) mapping [15] of don't cares in T_D' . According to the MTC mapping, the 'x' bits that follow a defined bit d (d can be either 0 or 1) are set to same value as d . This way the number of transitions in T_D' and, as a result, the average power consumption are minimized. Figure 5 demonstrates the two aforementioned 'x'-bit-filling approaches.

For assessing the power that will be dissipated during the scan-in of the two test sets of Figure 5, we make use of the Weighted Transition Metric (WTM) introduced in

[15]. According to the WTM, an indication of the power dissipated when a test vector v_j is shifted in the scan chain(s) of the CUT, is given by the sum of the propagation distances of v_j 's bit-transitions. More formally $WTM_j = \sum_{i=1}^{l-1} (l-i) \cdot (v_{j,i} \oplus v_{j,i+1})$, where l is the scan-chain length, $v_{j,i}$ are the bits of v_j and $v_{j,1}$ is the bit that is scanned-in first. In [15] it was shown that test vectors with higher values of WTM, dissipate more power. It can be easily calculated that the sum of the WTM of the vectors of Figure 5(a) (zero filling) is equal to 43, while the corresponding sum for Figure 5(b) (MTC mapping) is equal to 9 (we remind that the leftmost bit of the test vectors is shifted first in the scan-chain of the CUT).

| | | | | | | | |
|--|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| (a) zero filling of 'x' bits in T_D' | | | | | | | |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| (b) MTC mapping of 'x' bits in T_D' | | | | | | | |

Figure 5. (a) Zero filling and (b) MTC mapping of the 'x' bits of the example modified test set T_D'

As for the compression of T_D' , the zero filling approach could be effectively combined with some advanced run-length-code variant like the FDR code. However, the adoption of the MTC mapping will definitely create long run-lengths of 1s, which cannot be efficiently compressed by the FDR code. For that reason, another variant of the FDR, called alternating run-length, has been recently proposed by Chandra and Chakrabarty [14]. This code uses the same code words as the FDR for alternatively encoding run-lengths of 0s and 1s. It was shown experimentally in [14] that the alternating run-length code performs much better than the FDR, when MTC mapping is applied to the 'x' bits of the test set under compression. In fact, the combination "alternating run-length code – MTC mapping" may be more efficient in terms of test data compression even when compared against the "FDR code – zero filling" combination, since long run-lengths of 1s can be compressed as effectively as run-lengths of 0s. This assertion is verified by the example of Figure 5, where the compression of the first test set [Figure 5(a)] using FDR coding leads to a 40-bit set of encoded data, while for the second test set of Figure 5(b) the size of the alternating run-length-encoded data set is 38 bits. Therefore, we conclude that the combination of the MTC mapping with the alternating run-length code leads to much better scan-in power results with very small encoded data volumes.

One last comment that should be made is that the good results of the proposed method, which will be pre-

sented in the following section, are mainly due to the scan-chain inverter-insertion algorithm. If we just apply MTC mapping to the original test set of Figure 1 like the authors of [14], the sum of the WTM of the resulting test vectors will be equal to 52 and the alternating run-length-encoded data set will have a size of 50 bits. We remind that the corresponding values for the proposed technique are 9 (WTMs sum) and 38 (encoded data bits).

3. EVALUATION AND COMPARISONS

For evaluating the effectiveness of the proposed technique we implemented the algorithm of the previous section in C programming language for both the cases of maximizing run-lengths of 0s and 1s. Since all three, the proposed inverter-insertion criterion, MTC mapping and alternating run-length encoding, are extremely fast procedures, a CUT's test set can be optimized for run-lengths of 0s as well as for run-lengths of 1s, and the best result between those two can be chosen. In our experiments we considered the large ISCAS '89 benchmark circuits with full scan, assuming a single scan chain for each one. The test sets of the examined circuits were obtained by the Mintest ATPG program [12] with dynamic compaction. In Table 1 we present the results of the above-described experimental procedure.

Table 1. Experimental results of the proposed technique

| Circuit | No. of bits in T_D | 0s run-lengths maximizing | | | 1s run-lengths maximizing | | |
|---------|----------------------|---------------------------|----------|----------------------------------|---------------------------|----------|----------------------------------|
| | | Compression | | Scan-in avg. power (P_{avg}) | Compression | | Scan-in avg. power (P_{avg}) |
| | | No. of bits in T_E | Compr. % | | No. of bits in T_E | Compr. % | |
| s5378 | 23754 | 10604 | 55.36 | 2113 | 10094 | 57.51 | 2020 |
| s9234 | 39273 | 19996 | 49.08 | 3178 | 19432 | 50.52 | 3100 |
| s13207 | 165200 | 28628 | 82.67 | 6329 | 27328 | 83.46 | 5845 |
| s15850 | 76986 | 23272 | 69.77 | 11567 | 23066 | 70.04 | 11607 |
| s38417 | 164736 | 53094 | 67.77 | 93122 | 52610 | 68.06 | 92075 |
| s38584 | 199104 | 74250 | 62.71 | 85643 | 74250 | 62.71 | 85632 |

In the first two columns of Table 1 we provide the names of the examined benchmark circuits and the number of bits included in the original test sets (T_D). In columns 3 to 5 we present the compression and the scan-in average power results for the case of maximizing run-lengths of 0s, while in the following 3 columns the corresponding results for the case of maximizing run-lengths of 1s are shown. The average scan-in power is calculated as the mean WTM of the test cubes of each test set, i.e., $P_{avg} = (\sum_{j=1}^n WTM_j) / n$, where n is the test-cube volume.

As can be seen, the proposed technique leads to fairly good compression results, while the 1s run-lengths-optimizing approach is superior to the 0s run-lengths-optimizing, for almost all the benchmark circuits. Obviously, this is due to the characteristics of the circuits' test sets and certainly there will be cases in which the maximization of run-lengths of 0s will lead to better results. However, this is not a problem for the proposed technique, since, as mentioned above, thanks to the fast execution times, both cases can be examined.

We compare the proposed method against the approach of [14], which is the most recent technique that aims at the dual target of scan-in power and test-data volume minimization. Comparisons against techniques that deal with only one of these two problems (test-data volume or power dissipation) would not be fair, since, as explained in the introduction, the scan-power minimization objective does not allow full test-data-compression optimization and vice-versa. As expected though, the proposed method offers better compression results in all and in 4 out of 6 benchmark-circuit cases compared to the "zero filling-FDR encoding" technique of [2] and the selective Huffman approach of [3], respectively. Since the technique of [14] is based solely on the use of MTC mapping and the alternating run-length code, the reductions achieved by the proposed method demonstrate the effectiveness of the scan-chain inverter-insertion algorithm. The relative comparisons are given in Table 2. The best results of Table 1 were chosen for participating in the comparisons. We should note that exactly the same test sets were used for obtaining the results of both the compared techniques.

Table 2. Comparisons between the proposed technique and the approach of [14]

| Circuit | Test-data-volume comparisons | | | Scan-in power comparisons | | |
|---------|------------------------------|----------|-----------|---------------------------|----------|-----------|
| | [14] | Proposed | Reduct. % | [14] | Proposed | Reduct. % |
| s5378 | 11694 | 10094 | 13.68 | 2435 | 2020 | 17.04 |
| s9234 | 21612 | 19432 | 10.09 | 3466 | 3100 | 10.56 |
| s13207 | 32648 | 27328 | 16.30 | 7703 | 5845 | 24.12 |
| s15850 | 26306 | 23272 | 11.53 | 13381 | 11567 | 13.56 |
| s38417 | 64976 | 52610 | 19.03 | 112198 | 92075 | 17.94 |
| s38584 | 77372 | 74250 | 4.04 | 88298 | 85632 | 3.02 |

As can be seen from the comparisons of Table 2, the proposed scan-chain inverter-insertion approach significantly improves both the test-data compression and the scan-in power results in all cases. Compared to the technique of [14] and without any area or test-application-time penalty, the average gain is equal to 12.44% and 14.37% respectively.

A final remark that must be made concerns the test peak power, which is the maximum power dissipated in a single clock cycle during the testing process of a core. Although the proposed technique targets the reduction of average scan-in power, in our experiments we also counted the number of bit-transitions at every single clock cycle. We found out that the cycle in which peak-power dissipation occurs is, most of the times, the capture cycle. Thus, for reducing peak power, the proposed method can be combined with a technique that targets the minimization of capture-cycle transitions.

4. CONCLUSIONS

A novel approach that aims at the dual target of scan-in power and test-data volume minimization has been presented in this paper. This approach features a new scan-chain inverter-insertion algorithm that enables the effective modification of the original test set of a core to an-

other one with much better compression and scan-in power-dissipation characteristics. Furthermore, no performance or area penalty is imposed on the core under test, since the negated scan flip-flop outputs can be used for performing the necessary inversions. The proposed algorithm coupled with the minimum transition count mapping of don't cares in a test set and the alternating run-length code, offers significantly better test-data compression and scan-in power results than the corresponding works that have been already presented in the literature.

5. REFERENCES

- [1] A. Chandra and K. Chakrabarty, "System-on-a-chip test data compression and decompression architectures based on Golomb codes," *IEEE Trans. Computer-Aided Design*, vol. 20, no. 3, pp. 355-368, Mar. 2001.
- [2] A. Chandra and K. Chakrabarty, "Test data compression and test resource partitioning for system-on-a-chip using Frequency-Directed Run-Length (FDR) codes," *IEEE Trans. Computers*, vol. 52, no. 8, pp. 1076-1088, Aug. 2003.
- [3] A. Jas, J. Ghosh-Dastidar, Mom-Eng Ng and N. A. Touba, "An efficient test vector compression scheme using selective Huffman coding," *IEEE Trans. Computer-Aided Design*, vol. 22, no. 6, pp. 797-806, June 2003.
- [4] J. Rajski, J. Tyszer, M. Kassab and N. Mukherjee, "Embedded deterministic test," *IEEE Trans. Computer-Aided Design*, vol. 23, no. 5, pp. 776-792, May 2004.
- [5] I. Bayraktaroglu and A. Orailoglu, "Concurrent application of compaction and compression for test time and data volume reduction in scan designs," *IEEE Trans. Computers*, vol. 52, no. 11, pp. 1480-1489, Nov. 2003.
- [6] S. Wang and S. K. Gupta, "ATPG for heat dissipation minimization during scan testing," in Proc. ACM/IEEE Design Automation Conf., 1997, pp. 614-619.
- [7] P. Girard, L. Guiller, C. Landrault, S. Pravossoudovitch and H.-J. Wunderlich, "A modified clock scheme for a low power BIST test pattern generator," in Proc. VLSI Test Symp., 2001, pp. 306-311.
- [8] L. Xu, Y. Sun and H. Chen, "Scan solution for testing power and testing time," in Proc. Int. Test Conf., 2001, pp. 652-659.
- [9] J. Saxena, K. Butler, and L. Whetsel, "An analysis of power reduction techniques in scan testing," in Proc. Int. Test Conf., 2001, pp. 670-677.
- [10] V. Iyengar and K. Chakrabarty, "System-on-a-chip test scheduling with precedence relationships, preemption, and power constraints," *IEEE Trans. Computer-Aided Design*, vol. 21, no. 9, pp. 1088-1094, Sept. 2002.
- [11] O. Sinanoglu, I. Bayraktaroglu and A. Orailoglu, "Scan power reduction through test data transition frequency analysis," in Proc. Int. Test Conf., 2002, pp. 844-850.
- [12] I. Hamzaoglu and J. H. Patel, "Test set compaction algorithms for combinational circuits," in Proc. Int. Conf. Computer-Aided Design, 1998, pp. 283-289.
- [13] P. Rosinger, P. T. Gonciari, B. M. Al-Hashimi and N. Nicolici, "Simultaneous reduction in volume of test data and power dissipation for system-on-a-chip," *Electron. Lett.*, vol. 37, no. 24, pp. 1434-1436, Nov. 2001.
- [14] A. Chandra and K. Chakrabarty, "A unified approach to reduce SOC test data volume, scan power and testing time," *IEEE Trans. Computer-Aided Design*, vol. 22, no. 3, pp. 352-362, Mar. 2003.
- [15] R. Sankaralingam, R. R. Oruganti and N. A. Touba, "Static compaction techniques to control scan vector power dissipation," in Proc. VLSI Test Symp., 2000, pp. 35-40.