Rapid and Energy-Efficient Testing for Embedded Cores

Yinhe Han[†] **Yu Hu**[†] **Huawei Li**[†] **Xiaowei Li**[†] **Anshuman Chandra** [§]

[†] *Institute of Computing Technology, CAS, Beijing, 100080*

[†] *Graduate School of the CAS, Beijing, 100039*

yinhes@ict.ac.cn

[§] *Synopsys, Inc.*

Mountain View, CA 94043

Anshuman.Chandra@synopsys.com

Abstract

Conventional serial connection of internal scan chains brings the power and time penalty. A novel parallel core wrapper design (pCWD) approach is presented in this paper for reducing test power and test application time. The pCWD utilizes overlapping scan slices to reduce the number of scan slices loading. Experimental results on d695 of ITC2002 benchmark demonstrated that, about 2X shift time and 20X test power reduction can be achieved.

1. Introduction

Pre-designed intellectual property (IP) cores are increasingly used in System-on-a-chip (SOC) integration. Complex SOC integration creates some serious challenges for test. These challenges include test time, test data volume, testing power, and expensive tester, etc.[1]. Among these factors, test time is dominant in high-volume testing of ICs. Hence test time reduction is highly desirable.

Test time reduction can be obtained through scan chain reconfiguration and test data compression. Lee[2] proposes a methodology to use one scan-in signal to feed multiple circuits. The Illinois scan architectures in [3][4] show a broadcast architecture in which one scan-in signals are fed to multiple scan chains. [5] proposes a type of reconfigurable shared scan-in architecture. Using one scan-in to drive multiple scan chains can shorten scan chains, thus reducing the shift time of scan data. Test data compression is another way to reduce test application time. Golomb and VIHC codes have been presented in the work [6] and [7] to compress test data of embedded cores in SOCs. Reda[8] presents a mutation decoder to compress test data into data stream that indicates which bits need to be flipped in current test slice to obtain the subsequence one.

Another problem associated with testing SOCs is the high testing power. Testing of SOC cores in parallel poses a challenge that the parallelism possibly lead to exceed certain power thresholds which places the chip at a risk of damage. Reconfiguration of scan architecture can reduce testing power drastically. The utilization of externally controlled gates [9] and modification of scan chains through logic gate insertion between the scan cells [10][11] have shown the effectiveness of testing power reduction. [12] presents a 2D scan tree to shorten the scan chains to reduce the testing power. The modification on the hardware is at the price of performance degradation. On the other hand, some techniques [13] [14] don't modify the scan chains but adjust test patterns.

Reduction of test time and power can be achieved simultaneously [15][16] and [17]. However, these techniques are only suited for pre-design core providers, not integrators since they aim to reconfigure the scan architecture of embedded core.

The core wrapper design (CWD) is a DFT logic to provide test access for both core-internal testing, as well as core-external testing. The CWD commonly addresses two problems: transport capacity and width adaptation. Width adaptation is needed when TAM width and the number of TAM plugs are unmatched. When the width of TAM is insufficient for parallel loading, serialization is necessary at the input and output of ISCs [18] [19]. However, serialization will prolong the path of loading a bit into internal scan chains (ISCs), which multiplies the scan-in power and shift time as we investigated in the next section. To overcome the power problem, a novel parallel wrapper scan chains architecture is presented in this paper, which focuses on the connection of ISCs. Parallel connection can reduce testing power drastically. The high density of don't care bits(called as X-bit in this paper) in test data will be exploited to make consecutive test slices overlap, therefore reduce shift time and reduce testing power simultaneously.

The rest of this paper is organized as follows: In Section 2, we analyze the testing power and time penalty of serial connection. Section 3 will introduce the overlapping of scan slices. The main part of this paper: the parallel CWD is brought in Section 4. In the last section, some experimental results are reported to

[*] This paper is supported by the National Natural Science Foundation of China (NSFC) under grant No.90207002 and 60242001.

Proceedings of the 13th Asian Test Symposium (ATS 2004)
0-7695-2235-1/04 $20.00 © 2004 IEEE

show the effectiveness of the new wrapper architecture.

2. Power and Time Penalty of Serial Connection

Figure 1 shows a conventional serial connection in core wrapper design. Because available scan width of TAM is two, eight internal scan chains are connected into two wrapper scan chains(WSCs). The length of new WSCs is 4X longer than the ISCs in bare core. Prolonged scan path not only increase the test application time but also deteriorates the power dissipated in scan chains during shift.

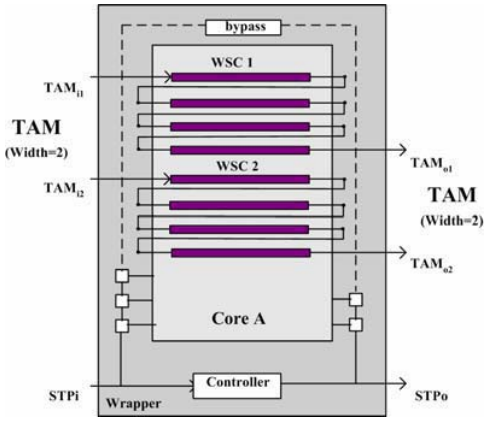

Figure 1. Conceptual architecture of serial connection in CWD

In this paper, *"Weight Transition Metric"(WTM)* [14] is used to estimate testing power dissipating in scan chains during scan test. We will limit ourselves to the scan-in power since the scan-out power has a similar property with scan-in power. We assume the length of all N ISCs is equal to l. The following property will give the relation between power consumption and width of TAM in wrapped core:

[Property 1] With the assumption that the transition probability of any two consecutive bits is same in serial connection, the N ISCs will be wrapped into M WSCs ($N>M$)it yields the following equation,

$$\frac{SIP^{w}_{core}}{SIP^{n}_{core}} = \frac{(N/M)\times l - 1}{l - 1} \approx \frac{N}{M}$$

Property 1 means that when N ISCs are wrapped into M WSCs, the testing power of wrapped core will become (N/M) times of the bare core.

Test application time(TAT) is also related to the width of TAM. The *TAT* for embedded core can be obtained by: $T=\{1+max(S_i, S_o)\}*P+min(S_i, S_o)\}$ [23], where P is the number of test patterns, and (S_i, S_o) denotes the length of the longest scan in(out) chains. [20] shows the TAT is also approximately inverse proportion with the width of TAM. However, as we investigated, there are numerous X-bit in test patterns. Since X-bit don't contribute to fault coverage, the shift time of them is redundant and necessary to reduce.

3. Slices Overlapping in Test Patterns

In *TAT*, the shift time *(ST)* is a main part. Shift time is the number of cycles needed to transfer the test data to internal scan chains. However, in test patterns, the main elements are don't care bits(X-bit). The observation of the X-bit distribution of ISCAS 89 circuits show that the X-bit in test patterns are "High density" and "Clustered". This distribution of X-bit can make two consecutive scan slices in test pattern to have a high probability to be overlapping. *Scan slices overlapping* means that scan slices are same. See the example in Figure 2, the pattern contains 10 scan slices. Each slice consists of 4 bits. The first and second slice in the original pattern are:{{1 X X 1},{X X 1 1}}. If the second and third bits in first slice and the first and second bits in the second slice are assigned to 1, they will be {{1 1 1 1}, {1 1 1 1}} and overlapping.

If consecutive slices are overlapping, a smart method to shift the slices to scan chains is that we load the first slice serially to an additional scan chain and then freeze this scan chain. In the following cycles, we only need to load the data in addition scan chains to internal scan chains in parallel. Thus, many redundant cycles can be saved. The scan architecture is looked like Figure 3. In Figure 3, four inputs of internal scan chains are parallel connected to one external scan chains (ESC). If we use the overlapping signature of test patterns, to the test patterns in Figure 2, only 3 slices need to be loaded serially, other slices can be loaded in parallel.

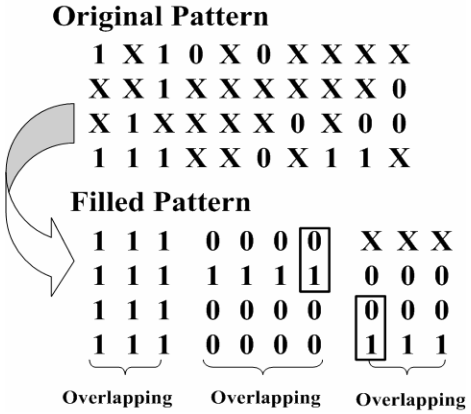

Figure 2. Slices overlapping in test pattern

4. Proposed Approach

4.1. Proposed Core Wrapper Design

In IEEE P1500 wrapper, the function inputs/outputs are wrapped by some scan cells. When width adaptation is needed, these scan cells (also called as *items*) and internal scan chains are serially connected into longer wrapper scan chains. As mentioned before, serial connection will prolong test time and deteriorate power consumption. In this section, we will propose parallel connection of internal scan chains overcome these two problems. In order to load the data in parallel, each scan-in and scan-out terminal should be also wrapped with a scan cell. The parallel CWD(pCWD) is shown in Figure 3. To stand out our proposed design, other hardware of IEEE P1500 has been omitted in the figure. Some XOR gates are inserted between the standard wrapper cells of scan out terminals. XOR gates and scan cell will be combined into some MISRs. All these MISRs are concatenated to form a large MISR. Large MISR can help us to reduce the error aliasing. The concatenated MISR is controlled by a normal clock: CLK. It is always active during the test application time. A simple logic unit: *U* is used to generate the clock of input external scan chains(iESCs) and ISCs. The logic test application of pCWD then proceeds by looping over two steps that typically consists of:

(1) Seed the iESCs: Firstly, the CLK1 is activated and CLK2 is frozen. Then the scan slices are shifted into iESCs from TAM. MISR in this time can be reconfigured into multiple scan chains. The results accumulated in the concatenated MISR are unloaded into TAM and transported to ATE to compare.

(2) Load the scan slices: the CLK1 is frozen and CLK2 is activated. Then the scan slices stored in the iESCs are loaded into ISCs in parallel. If multiple consecutive scan slices are overlapping, this step is repeated until a collision rises. In this procedure, the concatenated MISR is configured as a signature generator that accumulates and compact the contents of scan chains are shifted.

If the ISCs and iESCs are triggered by the different edges of a clock, the last cycle of step 1 and the first cycle of step 2 can be overlapped at a same scan clock. This condition can be easily satisfied by inserting a inverter in the clock tree.

The pCWD is fully compatible with IEEE P1500 standard. The clock: CLK may be the normal scan clock needed in the conventional serial connection. Compared with the standard architecture, an addition signal mode is needed to transform the states. It can be placed in TAM.

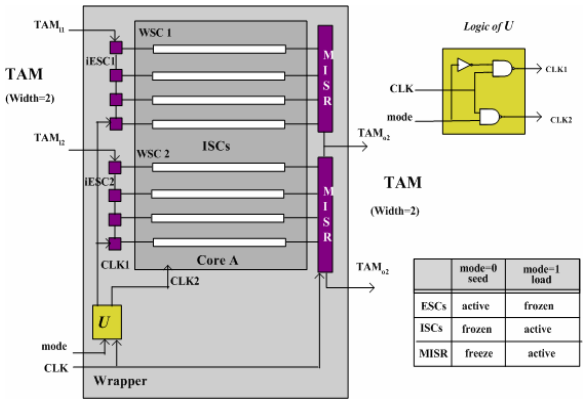

Figure 3. The proposed pCWD architecture

4.2. Testing Power Reduction and Test Pattern Process

As described in section 2, the testing power will increase with wrapping multiple ISCs into smaller WSCs. The pCWD reduce the testing power through shortening the shift path of a transition. The next property will show the fact that the testing power of pCWD is close to the testing power of bare core.

[Property 2] If the length of iESCs is far smaller than the length of ISCs, in parallel connection, testing power of wrapped core approximates to original test on bare core.

The *Property 2* shows that testing power dissipated in new *CWD* is independent with the width of TAM. The *pCWD* saves (*N/M-1*) times of scan in power dissipated in bare core compared with *sCWD*.

The test application time is determined by the number of overlapping blocks. To obtain the minimal number of blocks, a two-phase process: firstly partitioning test pattern to find a minimal number of blocks to cover all the slices, and then filling the left X-bit in blocks to keep overlapping. The minimum number of overlapping blocks can be found in a polynomial time since the index of scan slices must be ascending and continuous. In the following experiments, we use a Maximum Block Length Partition (*MBLP*) algorithm that finds the maximum block as early as possible in order to achieve this aim.

5. Experimental Results

Some experiments are conducted to verify efficiency of the proposed wrapper design method. The total source code in C language to implement the algorithm is about 3000 lines. The program is executed in a PC with a P4 1.6G processor and 384Mbits RAM.

The first series of experiments are performed on an academic SOC benchmark *d695* from Duke University

[20]. The specification of cores in this benchmark is listed in Table 1. The *Pat#*, *X-ratio* and *SC#* in Table 1 are the number of patterns, the ratio of X bits in total test pattern and the number of scan chains. These test patterns of all these cores are obtained from the MINTEST ATPG program [21].

Table 1. Test data for the six cores in d695

Cores	Circuits	*Pat#*	*X-ratio*	*SC#*
Core1	S9234	105	73.00%	4
Core2	S13207	234	93.20%	16
Core3	S15850	95	83.70%	16
Core4	S35932	12	38.60%	32
Core5	S38417	68	68.10%	32
Core6	S38584	110	82.30%	32

Table 2 shows the experimental results of d695. In this table, *ST#* is shift time and *SIP#* is scan in power. **w** is the input width of data TAM. We use *scan%* to measure the area overhead, where *scan%=(the number of scan cells in external scan chains) / (the number of scan cells in internal scan chains)*. In sCWD, the X-bit are random filled. To our pCWD, two phases process based on *MBLP* algorithm is used to optimize the overlapping. From the last column of Table 2, we can see about *2X* test time and *20X* testing power reduction can be obtained in pCWD with low area overhead. 20X power reduction in this case benefits from both scan chains reconfiguration and X-Fill algorithm. From the data of Table 1 and Table 2, we found that the test time reduction is related with ratio of X-bit and distribution of X-bit in test patterns. High density of X-bit can obtain better results. The density of X-bit in core 4 and core 5 is lower than other cores, so lower test time reduction is obtained.

The proposed architecture of pCWD can be a stand-alone time reduction technique. Table 3 will give some experimental results compared with previous techniques, such as Golomb code and mutation encoding. The test patterns are still generated by MinTest ATPG tool which are same as [6] and [8]. The data of both [6] and [8] are directly from [8]. The data of three techniques are measured by Shift Clock Cycles(same as SCC in [8]). In pCWD, the scan chains are divided into 8 scan chains and wrapped by one external scan chains. From the Table 3, our technique can get the higher time reduction in three of five cores. To total test time, our pCWD can achieve 11.7% improvement comparing with FDR code.

Table 2. Test Time and Power Reduction for cores in d695

Cores	w	sCWD		pCWD			Reduction	
		ST#	*SIP#*	*ST#*	*SIP#*	*Scan%*	*ST#*	*SIP#*
Core1	1	39680	2394242	18275	116163	2.02%	2.2X	20.6X
Core2	1	166144	28886996	33739	776552	4.57%	4.9X	37.2X
	2	83072	14362187	21283	407728		3.9X	35.2X
	4	41536	7182477	15055	201900		2.8X	35.6X
Core3	1	78624	11644669	28869	346435	5.23%	2.7X	33.6X
	2	39312	5828502	16093	205155		2.4X	28.4X
	4	19656	2890498	9705	139815		2.0X	20.7X
Core4	1	28672	4973024	13699	225754	3.63%	2.1X	22.0X
	2	14336	2568469	7091	119178		2.0X	21.6X
	4	7168	1280672	3787	66658		1.9X	19.2X
	8	3584	629784	2135	41474		1.7X	15.2X
Core5	1	166400	58704090	108089	1572643	3.85%	1.5X	37.3X
	2	83200	29098450	54985	926963		1.5X	31.4X
	4	41600	14380740	28433	602915		1.5X	23.9X
	8	20800	7265403	15157	441395		1.4X	16.5X
Core6	1	201664	71858550	86158	1770750	4.37%	2.3X	40.6X
	2	100832	35925936	44942	1017454		2.2X	35.3X
	4	50416	17927035	24334	645598		2.0X	27.8X
	8	25208	8943922	14030	459266		1.8X	19.5X

6. Conclusions

A new approach for core wrapper design based on parallel connection of internal scan chains has been proposed. It is compatible with IEEE P1500 and could be viewed as a partial extension of IEEE P1500 wrapper. Because the scan slices overlapping, pCWD can reduce test time since this architecture can load multiple useful bits into internal scan chains in one cycle. In order to obtain this reduction, a two phase process is presented to minimize the overlapping partitions The parallel connection will reduce testing power through shortening the path of a transition propagating. Furthermore, fill procedure in two phase process will reduce testing power in algorithmic level. If all the efforts are conducted, about *2X* test time reduction and about *20X* testing power reduction can be obtained compared to conventional serial connection.

7. Acknowledgment

The authors will thank Prof. Alex Orailoglu of University of California, San Diego for valuable discussion on related work of this paper when he had a travel to China. The first author will also thank Prof. Krishnendu Chakrabarty and Lei Li of Duke University for providing us with the test cubes used in [6].

Table 3. Test time Reduction for same pattern(MinTest)

Cores	PIs*Pat	Golomb [6]	FDR [22]	Mutation [8]	*pCWD*	
S13207	700*233	41658	20368	74423	31135	
S15850	611*94	92054	21590	26021	19399	
S35932	1763*12	59537	20946	7222	7020	
S38417	1664*68	92054	57066	45003	62583	
S38584	1464*110	104111	70328	73464	51895	*Imp#*
Total	503882	389414	190298	226133	170231	11.7%

Reference:

[1] ITRS. The International Technology *Roadmap for Semiconductors, 2001 Edition. http://public.itrs.net*.

[2] K. J. Lee, J. J. Chen, and C. H. Huang, "Using a single input to support multiple scan chains", *in Proceeding of IEEE/ACM International Conference on Computer-Aided Design*, pp. 74-78, 1998.

[3] I. Hamzaoglu and J. H. Patel, "Reducing Test Application Time for Full Scan Embedded Cores", *in Proceeding of International Symposium on Fault Tolerant Computing,* pp. 260-267, 1999.

[4] A. Pandey and J. H. Patel, "Reconfiguration Technique for Reducing Test Time and Test Data Volume in Illinois Scan Architecture Based Designs", *in Proceeding of VLSI Test Symposium*, pp. 9-15, 2002.

[5] S. Samaranayake, E. Gizdarski, N. Sitchinava, F. Neuveux, R. Kapur and T. W. Williams, "A Reconfigurable Shared Scan-in Architecture", *in Proceeding of VLSI Test Symposium, 2003.*

[6] A. Chandra, K. Chakrabarty, "System-on-a-Chip Test Data Compression and Decompression Architectures Based on Golomb Codes", *in IEEE Transactions on Computer-Aided Design,* 20:113-120, Mar, 2001.

[7] P. T. Gonciari, B. Al-Hashimi, and N. Nicolici, "Improving Compression Ratio, Area overhead, and Test Application Time in System-on-a-Chip Test Data Compression/Decompression", *in Proceeding of Design, Automation, and Test in Europe (DATE)*, 604-611, 2002.

[8] S. Reda, A. Orailoglu, "Reducing Test Application Time Through Test Data Mutation Encoding", *in Proceeding of Design, Automation and Test Conference in Europe (DATE),* pp.387-393, 2002.

[9] H. J. Wunderlich and S. Gerstendorfer,

"Minimized power consumption for scan based BIST", *in Proceeding of International Test Conference*, pp.85-94, 1999.

[10] O. Sinanoglu, I. Bayraktaroglu and A. Orailoglu, "Test Power Reduction Through Test Minimization of Scan Chain Transitions", *in Proceeding of VLSI Test Symposium,* pp. 166-171, 2002.

[11] O. Sinanoglu, A. Orailoglu, "Modeling Scan Chain Modifications For Scan-in Test Power Minimization", *in Proceeding of International Test Conference*, 2003.

[12] L. Xu, Y. Sun, H. Chen, "Scan solution for testing power and testing time", *in Proceeding of International Test Conference,* pp. 652-659, 2001.

[13] A. Chandra, K. Chakrabarty, "Combining Low-Power Scan Testing and Test Data Compression for System-on-a-Chip", *in Proceeding of IEEE/ACM Design Automation Conference,* pp. 166-169, 2001.

[14] R. Sankaralingam, R. R. Oruganti and N. A. Touba, "Static Compaction techniques to control scan vectors power dissipation", *in Proceeding of VLSI Test Symposium,* pp. 35-40, 2000.

[15] O. Sinanoglu, A. Orailoglu, "A Novel Scan Architecture for Power-Efficient Rapid Test", *in Proceeding of International Conference on Computer-Aided Design*, pp. 299-303, 2002.

[16] O. Sinanoglu, A. Orailoglu, "Test Data Manipulation Techniques for Energy-Frugal, Rapid Scan Test", *in Proceeding of Asian Test Symposium*, pp.202-207, 2003.

[17] D. Xiang, S. Gu, J. Sun, Y. Wu, "A Cost-Effective Scan Architecture for Scan Testing with Non-Scan Test Power and Test Application Cost", *in Proceeding of Design Automation Conference*, pp. 744-747, 2003.

[18] K. Chakrabarty, "Optimal Test Access Architectures for System-on-a-Chip", *in ACM Transaction Design Automation of Electronic Systems,* vol.6, No.1. pp.26-49, 2001.

[19] E. J. Marinissen, Y. Zorian, R. Kapur, T. Taylor and L. Whetsel, "Towards a Standard for Embedded Core Test: An Example", *in Proceeding of International Test Conference,* pp.616-627, 1999.

[20] V. Iyengar, K. Charkrabarty and E. J. Marinissen, "Co-Optimization of Test Wrapper and Test Access Architecture for Embedded Cores", *in Proceeding of International Test Conference*, pp.1023-1032, 2001.

[21] I. Hamzaoglu and J. H. Patel, "Test set compaction algorithms for combinational circuits", *in Proceeding of International Conference on CAD*, pp. 283-289, 1998.

[22] A. Chandra and K. Chakrabarty, "Test Resource Partitioning for SOCs", *IEEE Design&Test of Computers*, 18(5):80-91, 2001.

[23] E. J. Marinissen, S. K. Goel, and M. Lousberg, "Wrapper Design for Embedded Cores Test", *in Proceeding of International Test Conference*, 2000.