Partial Gating Optimization for Power Reduction During Test Application

Mohammed ElShoukry[1], C.P. Ravikumar[2], and Mohammad Tehranipoor[1]

[1]Dept. of CSEE, Univ. of Maryland Baltimore County, {elshoukry,tehrani}@umbc.edu

[2]ASIC Product Development Center, Texas Instruments India, ravikumar@ti.com

ABSTRACT

Power reduction during test application is important from the viewpoint of chip reliability and for obtaining correct test results. One of the ways to reduce scan test power is to block transitions from propagating from the outputs of scan cells through combinational logic. In order to accomplish this, some authors have proposed the setting of primary inputs to appropriate values or adding extra gates at the outputs of scan cells. In this paper, we point out the limitations of such full gating technique. We propose an alternate solution where a partial set of scan cells is gated. The subset of scan cells is selected to give maximum reduction in test power within a given area constraint. An alternate formulation of the problem is to treat maximum permitted test power and area overhead as constraints and achieve a test power that is within these limits using the fewest number of gated scan cells, thereby leading to least impact in area overhead. Our problem formulation also comprehends performance constraints and prevents the inclusion of gating points on critical paths. The area overhead is predictable and closely corresponds to the average power reduction.

I. INTRODUCTION

Power consumption during testing has become an important issue in modern day designs. It is increasingly higher than power during normal functional operation [1][2]. The low correlation between consecutively applied test patterns and their responses produces more transitions than in normal mode. Power consumption is especially important in today's chips, where larger numbers of transistors are packed into a smaller die size, higher frequencies are used, and aggressive timing requirements made at-speed testing methods highly important [3].

The abnormal power consumption during test can lead to adverse effects on the chip and the testing process, as outlined in [1][2][3]. The heat generated from high power dissipation can destroy the chip, cause reliability problems, or induce noise. Also IR and Ldi/dt drops on supply lines due to high switching currents may cause chips to falsely fail the test. With high power and energy consumption self testing of portable devices is rendered impractical.

For all these reasons, various techniques have been proposed to reduce the impact of the high power consumption during test application. The simplest of these are Ad hoc techniques such as slowing down the speed of the test clock, partitioning the circuit into blocks that are separately tested in a serial fashion, or providing extra packaging and cooling [2][3]. Since these will have negative impacts on test time and cost, and may not solve problems related to peak power, other methods were investigated. A good survey of many of these techniques can be found in [3].

For scan based designs in particular, switching activity in the combinational part contributes to a large portion of the total switching activity in the circuit [5]. These transitions are redundant during scan mode and are worth suppressing. A number of test power reduction techniques aim at reducing the number of transitions in the combinational part by blocking transitions occurring at scan cell outputs during scan mode. In [4] and [8], input control techniques have been suggested in which a certain assignment for primary inputs is selected in such a way as to block transitions at as many gates as possible, using an algorithm similar to the D-Algorithm in [4] and Kernighan-Lin (K-L) iterative improvement algorithm in [8].

A number of other techniques try to minimize transitions in the combinational part by gating the outputs of scan cells. A modified scan element has been suggested in [5]. By providing an extra gate at the output of each scan element, the output is held at a constant value during scan-in, and transitions in the scan flip-flops do not propagate to combinational logic. The gate is transparent in capture mode, and during normal operation. In [9] multiplexers have been used as gating elements, and in [7], a supply gating transistor has been introduced, so as to turn off the first level of logic connected to scan cell outputs while in scan mode.

In this paper, we show that with the proper selection of the percentage of gating elements, their position in the scan chain, and which output values to hold these gating elements at, we can control average power reduction to any desired level. In this partial gating approach, as opposed to full-gating approaches, we do not sacrifice as much area, and we can minimize the impact on delay. Often, maximum power reduction is not required in favor of area and performance, and also for the purpose of keeping the circuit under test in conditions similar to its operating conditions [10]. We show that for almost the same number of gating elements, the maximum achieved average power reduction can be more than twice the minimum reduction, for the same area overhead, only by varying the locations and the output of the gating elements. We describe an efficient way to evaluate how good a certain gating element placement is, and how closely this measure reflects the percentage of power reduction that can be achieved. Our proposed method is not computationally intensive and the area overhead is predictable. We show quantitatively that while gating scan cells can be a solution for average power, its impact on peak power may outweigh the advantage.

The paper is organized as follows. Section II provi

Proceedings of the 14th Asian Test Symposium (ATS '05)
1081-7735/05 $20.00 © 2005 IEEE

background and a general introduction to our approach. Section III describes our proposed method for partial gating. Section IV presents experimental results of implementing partial gating on several benchmark circuits. We provide our conclusions on the proposed method in Section V.

II. Background

In order to reduce or eliminate transitions in combinational logic, transitions should be prevented from propagating as close as possible to the outputs of the scan elements. For this purpose, some degree of controllability is needed at the gates which receive their input from scan cells. In [4] and [8], the authors proposed controlling primary inputs (PI) towards this purpose. These techniques however suffer from the low controllability provided by PI. Even for the smallest benchmark circuits we are using, the ratio of PI to the number of gates is 1.26%. Also PI may not always be directly settable as they may either be added to the scan chain or ignored completely on the expense of reduced fault coverage [6]. Techniques that rely on gating the scan cell outputs may be more effective in reducing power consumption.

As long as the gates at the first level of the combinational logic are controlled, the actual values at the outputs of the gates are immaterial. However, if only a partial subset of these gates are controlled to minimize overheads, the actual values on the outputs of the controlled gates are important. The addition of gating elements results in the following overheads: (a) Propagation delay in the gating element may change the critical path delays in the circuit (b) Gating elements result in area overhead, which can be significant if the number of scan cells is large. The selection of a subset of flip-flops for gating must be done with the aim of minimizing these overheads.

Inhibiting the transitions at the gate immediately following a scan flip-flop f does not guarantee that all transitions in the fanout cone of f will also be suppressed when partial gating strategy is chosen. Consider the example of Figure 1, where an OR gate is used at the output of the flip-flop and the output of the OR gate is controlled to logic 1 by setting PI = 1. The OR gate feeds a large number of AND gates, and potentially, all of them can toggle if the other inputs feeding the AND gates toggle during scan shift. On the other hand, if the output of the OR gate is controlled to 0 by setting PI = 0 and Q = 0, the toggling on the outputs of the AND gates can be prevented. The authors of [4] use a similar justification procedure for controlling power dissipation during scan test. However, their technique is to use only the PI for controllability, and we have explained earlier why this is not effective in most designs. We propose using a partial set of scan flip-flop outputs, which may be viewed as pseudo-inputs, to control the toggling activity in the combinational portion of the circuit. While this offers considerable flexibility, it also vastly increases the search space of possible solutions. In a circuit with p primary inputs and n scan flip-flops, it is easy to see that there are 2^{p+n} solutions when both PI and flip-flops are controllable, as opposed to 2^p solutions, when only the PI are controllable ($p << n$ in modern designs). The assignment that leads to lowest power is dependent on the circuit structure and must be carefully optimized.

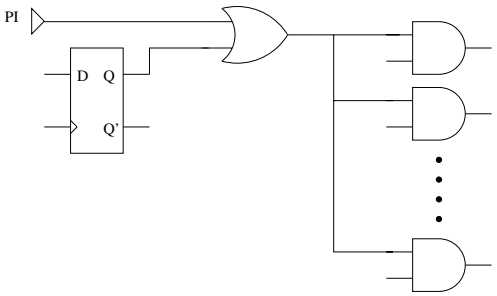

Figure 1. AND gates at the fanout of an OR gate.

III. Gating Point Selection and Optimization

A. Cost Function

As explained in the previous section, reaching an assignment of binary values to the PI and the pseudo-inputs to minimize the toggling activity in the combinational circuit is a difficult combinatorial optimization problem, even when some of the pseudo-inputs are not gated. To solve this problem through heuristic techniques, one needs a cost function to compare one assignment against another. An exact measure will involve counting the number of toggles in the combinational circuit, but the computational complexity of implementing such a metric is high, since it will involve logic simulation of the circuit for every pattern.

We use a measure that attaches a weight to the number of gates that are guaranteed *not to switch* during the process of scanning in a pattern. An X at a pseudo-input corresponds to a non-gated scan flip-flop. Gated scan cells will have 0 or 1 at their output. After logic simulation, any gate that has X at its output is assumed to be changing as scan cells change. The weight attached to a gate g that satisfies this criterion is the fanout of g. In other words, our cost function for a pattern is of the form

$$Cost = \sum_{g|output(g)=1 \parallel output(g)=0} fanout(g) \qquad (1)$$

Ideally, the weight can include the exact capacitance driven by the output of g. The capacitance information is not available until later in the design flow, and the fanout of the gate output can be taken as a measure of this capacitance. A simple static technique to compute the cost function is shown below. The cost function is a measure of the power saving resulting from partial gating of the scan flip-flops for a scan vector V.

COMPUTE_COST_FUNCTION (V)
/* V is a scan pattern that assigns 0, 1, or don't-care values X to PI and pseudo-inputs.

1) $cost = 0$
2) Perform a 3-valued logic simulation for vector V
3) **for each** gate g **do**
   a) **if** output (g) = 1 **or** output(g) = 0 **then** $cost = cost +$ fanout(g)
4) **return** $cost$

The three-valued logic simulation used in cost estimation has its limitations, although it is fast. Logic simulation doesn't distinguish between X and $\overline{X}$. If both appear at an OR gat

for example, $X+\overline{X}$ will evaluate to X instead of 1. Using a logic value such as $\overline{X}$ in the simulation will correctly predict the output at the OR gate. Thus, the procedure may underestimate the power saving resulting from partial gating. Similarly, the procedure does not perform a delay simulation of the gates and can hence not predict any glitching. Ignoring glitching can result in overestimation of power saving. Because of these two opposite effects, the estimate of cost function can be expected to be reasonably accurate when comparing one vector to another.

B. PI and Gating Assignment

Formally, the input assignment problem is defined as follows. Given a full scan circuit with p primary inputs, n scan flip-flops, and a certain area overhead given as a certain percentage of the scan cells, we wish to identify the subset of flip-flops whose outputs must be gated such that the saving in scan test power is maximum while constraining the area and performance overhead due to gating elements. Any one of the gating schemes discussed in [5], [7], or [9] can be used. An assignment algorithm based on random search is shown below. The algorithm repeatedly generates random vectors of size $p+n$ and samples a large population of the total search space of 2^{p+n} vectors. The bit $V[i]$ is set to an X with a probability $probX_i$ that is generally equal to (1 - overhead percentage). The generated vector is evaluated for its power saving metric using the cost function of the previous subsection. A lower value for probability $probX_i$ implies a higher probability of inserting a gating element at scan flip-flop i. In this way we can control the area overhead.

INPUT_ASSIGNMENT

1) Remove the scan flip-flops and convert them into pseudo-inputs
2) $MaxResult = -1$;
3) $MinResult = \infty$;
4) **for** k = 1 **to** *Number of Iterations* **begin**
   a) /* Generate a random vector $V[1 : p+n]$ of $p+n$ bits */
   b) **for** i = 1 **to** $p+n$ **do**
      i) Set $V[i]$ to X with probability $probX_i$;
   c) $Result$ =COMPUTE_COST_FUNCTION(V);
   d) **if** $Result > MaxResult$ **then** $BestVector = V$;
   e) **if** $Result < MaxResult$ **then** $WorstVector = V$;
5) **end**
6) **return** *BestVector*, *WorstVector*

C. Don't Care Probability and Critical Paths

$probX_i$ can be either set as a single value for all inputs or it can be controlled independently for each input, forming a probability vector. In this way we can control the placement of gating elements at certain points. For example, for scan elements on the critical paths we can assign don't care probabilities of 1. In such a case the algorithms generating bit pattern will always place an X at this bit position and never requires a gating element.

IV. Experimental Results

In order to evaluate the proposed method, we performed experiments on several ISCAS'89 benchmark circuits. We used NOR and OR gates for gating to 0 and 1, respectively. Table I shows the reduction in average and peak power when all the scan flip-flops are gated. Column 2 shows the average power reduction when gating overhead is excluded, which makes it independent of the type of gating used. It also shows a high contribution of the power in scan chains in the total circuit power, since the reduction is 38% in the largest circuit (s38584) even when the combinational part is totally turned off. This can be a direct result of the particular library we used (such as the high clock pin input capacitance). Layout optimizations and buffering can help reduce this power, independent of the use or non-use of gating, and the type of gating elements. Column 3 shows the average power reduction when power in the gating elements is considered, for our particular NOR/OR gates used. It shows that significant power is consumed in the gating elements themselves, especially when the number of gating elements is large, as in full-gating. In all of the benchmarks, the peak power increased, from about 5% and up to 60% when the gating elements overhead is considered. This increase is due to a large number of gating elements changing state either when the scan chains change from shift mode to capture mode, or during capture itself.

Tables II and III show the results when the number of gating elements was 50% and 80% of the total number of scan flip-flops. In each table, we show the reduction in average power and peak power when (a) best vector was used and (b) worst vector was used. The best and worst vectors were found using the INPUT_ASSIGNMENT procedure explained in the previous section. The number of iterations of random search was set to 10000. The entries under %ckt column are the power reductions in the entire circuit, whereas the entries under %comb are the power reductions in the combinational part only, and %(comb+GE) is the power reduction in both combinational logic and the gating elements. In Table I (full-gating), we have not included a column for %comb since there is no switching activity in the combinational part during scan-in. %comb serves to give a measure that is independent of both the power in the scan chains, and the type of gating element used. As discussed before, both can be subject to further optimizations.

We make the following observations.

- For 50% gating, the ratio of the average power reduction in the best and worst vectors ranges from 1.29 to 2.47. The difference tends to narrow as the gating percentage is increased as power in the gating elements starts to dominate.
- When 50% of the flip-flops are gated, the achieved reduction in average power is more than 50% of the reduction when all the flip-flops are gated, even when gating overhead is included if gating elements are carefully selected.
- If we compare Column 3 in Table I, and Column 2 in Table

TABLE I

POWER REDUCTION USING FULL GATING

Benchmark	%ckt avg. w/o GE	%ckt avg. with GE	%ckt peak w/o GE	%ckt peak with GE
s5378	50.72%	38.49%	14.00%	-5.67%
s9234	55.89%	45.44%	7.74%	-4.74%
s13207	43.15%	28.96%	1.87%	-17.89%
s15850	42.63%	28.42%	-8.86%	-30.98%
s35932	42.44%	30.11%	-35.52%	-42.70%
s38417	34.97%	21.45%	6.70%	-17.41%
s38584	38.83%	24.15%	-31.50%	-60.03%

TABLE II

AVERAGE POWER AND PEAK POWER REDUCTION FOR 50% GATING ELEMENTS

Benchmark	Average Power								Peak Power (with GE)	
	Best Vector (B)			Worst Vector (W)					Best	Worst
	%ckt	%comb	%(comb+GE)	%ckt	%comb	%(comb+GE)	B/W	%(B-W)	%ckt	%ckt
s5378	30.13%	67.52%	59.42%	12.18%	28.48%	24.02%	2.47	35.40%	10.47%	5.87%
s9234	31.15%	61.82%	55.73%	14.61%	29.92%	26.14%	2.13	29.60%	5.69%	7.96%
s13207	23.92%	65.18%	55.43%	10.20%	32.44%	23.65%	2.34	31.78%	-2.94%	-4.56%
s15850	22.52%	62.60%	52.82%	14.95%	43.96%	35.07%	1.51	17.75%	-3.69%	-6.36%
s35932	20.01%	55.79%	47.15%	14.36%	41.35%	33.84%	1.39	13.31%	-20.34%	-11.15%
s38417	15.65%	56.30%	44.75%	10.07%	38.66%	28.78%	1.55	15.97%	-4.96%	-4.09%
s38584	16.49%	53.10%	42.48%	12.74%	42.86%	32.81%	1.29	9.67%	-24.94%	-20.90%

TABLE III

AVERAGE POWER AND PEAK POWER REDUCTION FOR 80% GATING ELEMENTS

Benchmark	Average Power								Peak Power (with GE)	
	Best Vector (B)			Worst Vector (W)					Best	Worst
	%ckt	%comb	%(comb+GE)	%ckt	%comb	%(comb+GE)	B/W	%(B-W)	%ckt	%ckt
s5378	39.22%	87.81%	77.34%	23.63%	57.23%	46.59%	1.66	30.76%	5.71%	5.23%
s9234	44.83%	89.69%	80.20%	33.26%	67.91%	59.51%	1.35	20.69%	1.56%	2.66%
s13207	32.05%	89.35%	74.28%	19.31%	58.63%	44.74%	1.66	29.54%	-12.11%	-6.87%
s15850	28.26%	80.47%	66.29%	27.46%	77.98%	64.42%	1.03	1.86%	-14.47%	-10.05%
s35932	29.75%	83.12%	70.09%	26.62%	74.91%	62.73%	1.12	7.36%	-33.79%	-25.00%
s38417	23.95%	85.38%	68.49%	17.27%	65.97%	49.37%	1.39	19.12%	-9.98%	-8.37%
s38584	25.92%	83.80%	66.76%	22.09%	72.20%	56.89%	1.17	9.86%	-40.35%	-32.95%

III, we can see that when gating overhead is included, the total reduction in average power is almost the same as the reduction when full gating is used. This shows that with less area overhead we can achieve almost the same average power reduction and have less impact on peak power. Comparison of Columns 5 in Table I, and Column 10 in Table III shows that while peak power increases in both cases, the effect of full gating is worse.

- Comparing Tables II and III, we find that the average power reduction in the combinational part always exceeds the gating percentage if gating points are properly selected, but when gating overhead is included, the gating overhead starts to offset the saving achieved by extra gating. As the gating percentage is increased, the power in the gating elements themselves goes higher.
- Gating overhead can have considerable impact on both average and peak power, which suggests the use of special types of gates as gating elements or the use of low overhead gating techniques, such as the one in [7], that will reduce their impact.

These observations clearly indicate that careful selection of gating elements can prove to be very effective in test power reduction, and the optimization of the gating elements themselves is necessary to reduce their impact on power saving.

V. CONCLUSION

Other than scan data volume reduction, which has received a lot of attention in the recent past, another issue that has generated a great deal of interest is scan test power reduction. Full-gating techniques that authors have proposed in the past, to convert all the scan flip-flops to gated scan flip-flops, is area-intensive, complex from a physical design viewpoint, and intrusive from the timing closure viewpoint. The main contribution of this paper is to take advantage of the structural properties of the circuit to avoid full gating. A user-specified percentage of scan cells can be converted to gated scan cells in the proposed methodology, allowing the user to control the overheads of scan gating. Heuristic algorithms were used to select scan cells for gating. In our results, even 50% gating gave up to 30% average power reduction where a 100% gating provides about 38% reduction. We showed that the overhead in the gating elements themselves can have serious effects and cannot be ignored. We attempted to show results that are independent of the power consumed in the scan chain or the type of gating elements used.

REFERENCES

[1] Y. Zorian, "A distributed BIST control scheme for complex VLSI devices," in Proc. *VLSI Test Symp. (VTS'93)*, pp. 4-9, 1993.

[2] S. Wang and S.K. Gupta, "ATPG for heat dissipation minimization during test application," in Proc. *Intl. Test Conf. (ITC'94)*, pp. 250-258, 1994.

[3] P. Girard, "Survey of low-power testing of VLSI circuits," *IEEE Design & Test of Computers*, vol. 19, no. 3, pp. 80-90, 2002.

[4] T. Huang and K. Lee, "Reduction of power consumption in scan-based circuits during test application by an input control technique," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 20, no. 7, pp. 911-917, 2001.

[5] S. Gerstendorfer and H.-J. Wunderlich, "Minimized power consumption for scan-based BIST," in Proc. *Intl. Test Conf. (ITC'99)*, pp. 77-84, 1999.

[6] L. Whetsel, "Adapting scan architectures for low power operation," in Proc. *Intl. Test Conf. (ITC'00)*, pp. 863-872, 2000.

[7] S. Bhunia, H. Mahmoodi, S. Mukhopadhyay, D. Ghosh, and K. Roy, "A Novel Low-Power Scan Design Technique Using Supply Gating," in Proc. *Intl. Conf. on Computer Design: VLSI in Computers and Processors (ICCD'04)*, pp. 60-65, 2004.

[8] Seongmoon Wang and S.K. Gupta, "ATPG for Heat Dissipation Minimization during Scan Testing," in Proc. *Design Automation Conf. (DAC'97)*, pp. 614-619, 1997.

[9] Xiaodong Zhang and K. Roy, "Power reduction in test-per-scan BIST," *IEEE On-Line Testing Workshop*, pp. 133-138, 2000.

[10] K.M. Butler, J. Saxena, A. Jain, T. Fryars, J. Lewis, and G. Hetherington, "Minimizing Power Consumption in Scan Testing: Pattern Generation and DFT Techniques," in Proc. *Intl. Test Conf. (ITC'04)*, pp. 355-364, 2004.