

A Clock-gating Based Capture Power Droop Reduction Methodology for At-Speed Scan Testing¹

Bo Yang[‡] Amit Sanghani[‡] Shantanu Sarangi[†] and Chunsheng Liu[§]
[‡]DFT Engineering, NVIDIA Corp. [†]AMD Corp. [§] Test Development, Altera Corp.
2701 San Tomas Expressway One AMD Place 101 Innovation Drive
Santa Clara, CA 95050, USA Sunnyvale, CA 94088, USA San Jose, CA 95134
byang, asanghani@nvidia.com shantanu.sarangi@amd.com cliu@altera.com

Abstract

Excessive power dissipation caused by large amount of switching activities has been a major issue in scan-based testing. For large designs, the excessive switching activities during launch cycle can cause severe power droop, which cannot be recovered before capture cycle, rendering the at-speed scan testing more susceptible to the power droop. In this paper, we present a methodology to avoid power droop during scan capture without compromising at-speed test coverage. It is based on the use of a low area overhead hardware controller to control the clock gates. The methodology is ATPG (Automatic Test Pattern Generation)-independent, hence pattern generation time is not affected and pattern manipulation is not required. The effectiveness of this technique is demonstrated on several industrial designs.

1 Introduction

Scan-based Design-for-Test (DFT) technique is a common practice in integrated circuit (IC) manufacture test [12]. A standard scan operation includes two procedures: shift and capture. In both procedures, power dissipation is usually several times higher than that in normal operation because scan test intends to exercise the chip as much as possible. Excessive power dissipation can cause various problems such as overheating, increased delay and noise, IR drop etc. [2, 4, 6, 7, 8, 9, 10].

Although it is an easy work-around to reduce power during shift by decreasing scan frequency, such method causes increased test time hence higher production cost. Moreover, similar compromise cannot be used during at speed capture. For most state-of-the-art designs, 40nm or deeper processing is becoming increasingly popular, where delay-related

failures are becoming a dominant type of defects. Therefore, at-speed structural test is necessary. To ensure the test quality, circuits during scan capture procedure must be clocked at normal functional speed. This requirement exacerbates the problem and necessitates a proper scan scheme that can alleviate power dissipation without compromising the at-speed nature of the test.

For modern large chips such as GPU (Graphic Processing Unit), the power issue is more challenging because such designs demand astounding power when they are loaded, due to their scale and the nature of parallel processing. E.g. latest GPU chips consist of billions of transistors and can easily consume a couple of hundred watts when fully loaded. Such an amount of power in normal operation mode will climb up by another 4X or 5X during scan test. This essentially exceeds the capability of the chip's power grid, or even the power supply, causing significant power droop across the die.

Figure 1 shows a power droop phenomenon on a latest GPU chip during at-speed testing, where the launch-off-capture clocking scheme is used [12] for scan capture. I.e., during scan capture, a launch clock cycle will be first emitted followed by a capture clock cycle. The wave is captured by ATE (Automatic Test Equipment) VDD sensor. It can be seen when launch clock is fired, there is a significant drop (0.525v) on VDD followed by a series of oscillations. The VDD does not have enough time to recover before the capture clock is fired (933ps between two clock pulses), causing massive failures. Such a severe power droop invalidates scan-based at-speed test results. Therefore, without constraint on power dissipation, it is difficult to conduct scan-based at-speed tests on the current power-hungry designs.

Enormous prior work has been conducted to target the power reduction in scan test [11]. Novel scan architectures with power reduction capability during scan are presented in [1, 4]. ATPG improvements are proposed in [3, 13], where power is reduced by manipulating specific bits in scan patterns. Smart voltage scaling [5] and specific clock schemes

¹The work of C. Liu and S. Sarangi was conducted when they were with NVIDIA Corp.



Figure 1. VDD power droop observed on ATE.

[6] can also be combined with scan procedures to either reduce both dynamic and leakage power, or control power droops during capture. With power droop being an increasingly critical issue, various ATPG methods have attempted to detect faults caused by power droop for increased coverage [7, 9, 10].

Most of these earlier methods require either changes in scan architecture, or involvement of power-aware ATPG, or a combination of the two. These requirements have limited their applications in many industrial designs. Any change made to scan architecture will need an overhaul to the entire scan flow and enormous verification efforts. Meanwhile, many design companies rely on commercial ATPG tools for at-speed pattern generation, hence they do not have the flexibility to customize the pattern generation flow. Tight time-to-market has made such practices even more infeasible. Another possibility is to use power-aware features in existing commercial ATPG tools. However, our experiences have shown that such features are still far from practical for large designs because of the prohibitive run time. As a result, we desire a novel scheme for mitigating power droop in scan-based at-speed testing with the following features:

1. Compatible with existing standard scan architecture and DFT flow;
2. No ATPG pattern manipulations needed;
3. Hardware overhead must be minimum;
4. No loss on coverage, no significant increase in pattern count;
5. Produce deterministic activities within power budget.

These issues have not been addressed simultaneously in any previous work. In this paper, we present a method that can mitigate the capture power droop during at-speed scan test while meeting the above requirements. This technique partitions clock gates and the downstream flip-flops into groups and uses a novel controller to constrain the transition activities during capture within the power budget. In

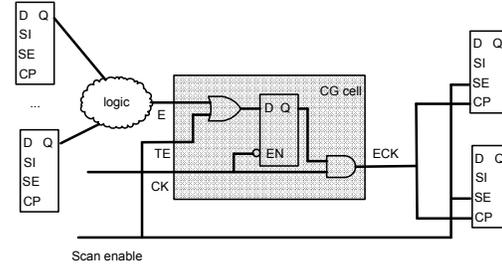


Figure 2. Clock gating structure.

section 2, we present the proposed low capture power architecture. In section 3, we describe the flow to group the clock gates to meet the power budget requirement. Section 4 presents experimental results on some latest industrial designs. And finally Section 5 concludes the paper.

2 Clock gate-based low capture power architecture

2.1 Motivation

Clock gating has been widely employed as one of the most popular dynamic power-saving techniques in synchronous circuits. Gating logic is added to the clock tree so that the downstream clock can be disabled and flops will not switch, eliminating dynamic power consumption. A representative clock gating design using a RZ(return to zero) clock is shown in Figure 2. An integrated clock gating (CG) cell, shown as the shaded box, has a functional enable port E and a test enable port TE. Generally TE port is connected to scan enable to make sure shift clock is not disturbed by functional logic. The logic driving port E is determined by functional operation. In scan capture, ATPG tool will attempt to set the E pin to 1 to enable the clock ECK so that the downstream flops can capture. If the logic in the input cone of E pin is hard for ATPG to handle, more patterns are needed to achieve the coverage.

In coverage-critical pattern generation, ATPG tools tend to set the E cone logic to enable as many CGs as possible for better coverage. As a result, more flops are toggled in scan capture mode than in full-load functional mode, causing prohibitive dynamic power dissipation and potential power droop, which can lead to yield loss. Moreover, since the power grid is usually designed to support only functional power budget, it may not be able to sustain an excessive scan capture power, in which case at-speed testing can not be performed at all. Most ATPG based power reduction techniques [3, 13] rely on the manipulation of specific bits in scan patterns, so that only a small amount of CGs are enabled per pattern. Such methods usually compromise test coverage for a given number of patterns, or lead to more patterns to achieve a given coverage. It is possible that within

Table 1. Inefficiency of Power Aware ATPG.

ATPG	Regular	Power-aware
Flop count	7.5M	
CGed flops	87.7%	
power budget	NA	15%
pattern count	100	
coverage	62.72%	38.08%
run time(sec)	36261.03	1724167.88

the ATE’s pattern limit, such power-aware ATPG can not achieve the desired coverage.

In order to meet the power budget, the power-aware ATPG tools need to apply a large number of constrains on scan cells, hence many patterns are dropped. Such repetitive attempts will cause prohibitive pattern generation time, rendering many of such power-aware ATPG methods impractical for large designs. To illustrate the inefficiency of such methods, we list in Table 1 some data collected from pattern generation on a GPU design using a commercial ATPG tool with power-aware option turned on.

The design has 7.5M flops and 87.7% of them are clock gated. It takes about 10 hours to generate 100 patterns for transition faults with a coverage of 62.72% without power-aware option. When power-aware is enabled, however, it takes 20 days to generate 100 patterns on a power budget of 15% of the full power level. I.e. the ATPG tool attempts to control the E pin of the CGs so that no more than 15% of flops are toggled during capture of each pattern. The data clearly show that it is infeasible for such power-aware ATPG tool to meet all requirements of test coverage, power budget, and time-to-market. Note that results from other ATPG tool could vary.

2.2 Low capture power scheme using transition controller based on one-hot decoder

To address the power droop issue in scan capture, we propose a scheme that relies on the use of a novel hardware controller to provide deterministic power control. It is based on a one-hot decoder through which ATPG tool can selectively control the E pin of the clock gates, hence control the amount of transition activities of flops (i.e. power level, since capture power is proportional to the number of transitions in flops). The allowed power budget is programmable through JTAG register and the scheme does not require any manipulation on ATPG tool.

A simplified scheme is shown in Figure 3. The shaded boxes represent the clock gate cell shown in Figure 2. An AND gate is added in front of the E pin of every CG. One input of the AND gate is the original functional enable logic, the other is driven by a one-hot decoder based transition controller logic (dotted box). We refer to this leg of the AND gate as LPE, or "Low Power Enable". Note that the

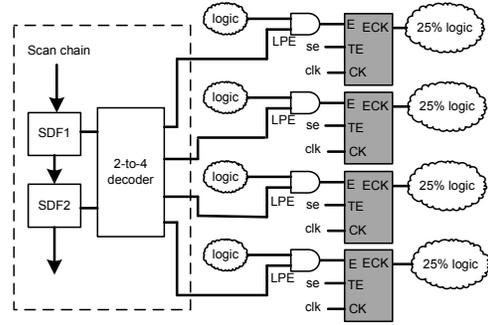


Figure 3. A simplified low capture power scheme using a 2-to-4 decoder.

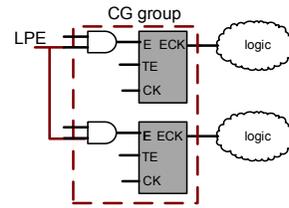


Figure 4. A clock gate group with two CGs.

addition of the AND gate adds extra delay on the enable path. However our experience shows that the effect can be minimized through better timing effort. And the AND gate can be built into the CG as an integrated low power CG cell to achieve better timing. For the sake of simplicity, we assume a fixed power budget of 25% is required, meaning that in any pattern no more than 25% flops can be clocked during capture. We further assume that the design is small so that all flops can be covered under 4 CGs, and each CG will cover a similar number of flops. Then we only need to insert a 2-bit controller. The LPE signal of each CG is driven by one output of the controller. The controller in its simplest case can be a 2-bit scan flops and a 2-to-4 one-hot decoder as show in the figure. No matter what values ATPG tool loads into flops SDF1 and SDF2, one and only one of the 4 outputs will be active high and only one CG can pass capture clock. This guarantees that a 25% power budget is never violated. Note that 25% is an upper bound but ATPG tool will attempt to toggle most flops controlled by the enabled CG. To achieve a finer power level control, various decoders can be used. E.g., a 4-to-16 decoder is able to provide a power level granularity of 6.25% of the full power.

In a large design, there can be thousands of CGs. Therefore, we cannot control each individual CG but need to organize them into a small number of CG groups for easy control. The total number of flops covered under each CG group, i.e. the load of each CG group, is similar. This organization will be described later in Section 3. All CGs in a CG group share an LPE signal, as shown in Figure 4.

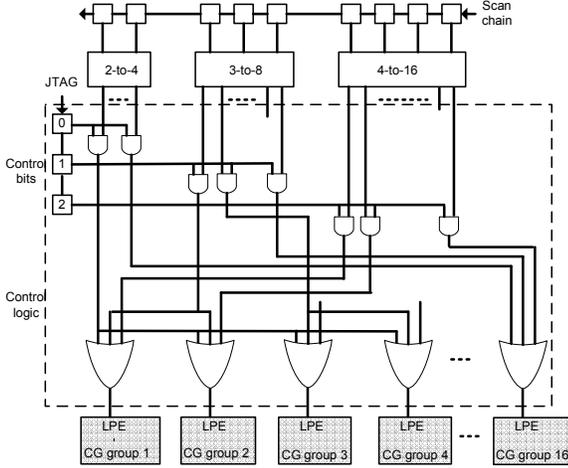


Figure 5. A generic controller.

This simplest transition controller can only provide a fixed power budget because it only includes one decoder. Moreover, it can only issue capture clock to flops in one CG group in a pattern. If there exists a path involving two flops in two different CG groups, e.g. a cross-domain path, coverage is lost because such a path can never be exercised. To address these issues, we will next describe a more practical transition controller.

2.3 Controller design

Figure 5 shows a generic controller. We refer to this as "generic" because it can represent a set of designs that can consist of an arbitrary combination of different decoders and control logic. For the sake of simplicity and illustration, in this figure we assume the clock gates in the partition are organized into 16 CG groups, and each group has similar load of scan flops. We further assume that only 2-to-4, 3-to-8 and 4-to-16 decoders are available and we want the power level (the percentage of switching flops during capture) to be adjustable upto 50% of full power at a granularity of 6% (or $\frac{1}{16}$). Although there are various possible designs, we present a simple case in which only one decoder of each type is used.

As explained earlier, the inputs of the decoders are from scan chains so that ATPG tool can determine the value on each input. The dotted box represents a control logic with several JTAG registers as control bits. The values of these bits can be set through JTAG for a desired power level. Since it can represent a large number of specific designs, in this figure we illustrate only one example in which three control bits are used. Each control bit is associated with one decoder and determines if the decoder can affect the control of the CGs (=1), or not (=0).

In this specific example, we first partition the 16 CG groups into 4 groups, i.e. $\{1,2,3,4\}$, $\{5,6,7,8\}$,

$\{9,10,11,12\}$, $\{13,14,15,16\}$. Each group will be driven by one output of the 2-to-4 decoder. Similarly, the CG groups can be organized in 8 groups, i.e. $\{1,2\}$, $\{3,4\}$, $\{5,6\}$, $\{7,8\}$, $\{9,10\}$, $\{11,12\}$, $\{13,14\}$, $\{15,16\}$, and each group can be driven from one output of the 3-to-8 decoder. For the 4-to-16 decoder, each output can drive a CG group. The outputs from each decoder are then ORed to drive the LPE signal of the CG group.

It can be seen that this design effectively implements several important functions. First, it doesn't change the existing scan architecture and doesn't involve pattern manipulation. Second, any clock gate is controllable. The JTAG control bits will first select what decoder can be effective, and then ATPG tool will determine which CG group should be turned on/off in a specific pattern. Third, any two CG groups can be turned on simultaneously. If a single decoder is enabled (through control bits), then only the clock gates in one CG group can be turned on. However if two or more decoders are enabled, then any two CG groups can be turned on in a pattern, hence it becomes possible for ATPG tool to target a path between flops in two CG groups. E.g., if a path to be covered involves two flops in CG groups 12 and 13, respectively, the scheme can provide several possibilities. One possibility is to enable CG groups $\{9,10,11,12\}$ through 2-to-4 decoder and $\{13,14\}$ through 3-to-8 decoder simultaneously, another is to enable $\{11,12\}$ through 3-to-8 decoder and $\{13\}$ through 4-to-16 decoder, etc., depending on ATPG tool.

Finally, a desired power level can be set through the control bits. Each group from the 2-to-4 decoder can yield a maximum power level of 25%, and the other two decoders can yield 12.5% and 6.25%, respectively. Therefore, we can obtain a maximum power level of any combination of these numbers. For granularity of 6.25%, the 4-to-16 decoder must be enabled. We should note again that these power values are proportional to the number of flops in the CG groups that are enabled, and hence they represent the maximum power level because even if a CG group is enabled, ATPG tool may not be able to toggle ALL the flops in it in a pattern. Therefore, this approach is based on a pessimistic expectation.

Also note that for the control logic example shown here, it is possible that the two groups enabled by two decoders are overlapped, e.g. $\{15,16\}$ from the 3-to-8 and $\{16\}$ from the 4-to-16. Hence the maximum power level from this setup will have a lower bound of 12.5% and an upper bound of 18.75% (6.25+12.5). Obviously, one can design a more complex control logic to avoid such overlap if necessary. A partial truth table of the input (control bits) and the output (the power level) is shown in Table 2. Only representative control bits values are listed.

Another observation on this scheme is that if ATPG tool needs to toggle a scan flop during capture, it will attempt

Table 2. Partial truth table for example control logic in Figure 5.

Control bits	Power level(%)	
	Min	Max
000	0	0
001	25%	25%
010	12.5%	12.5%
011	25%	37.5%
...		
111	25%	43.75%

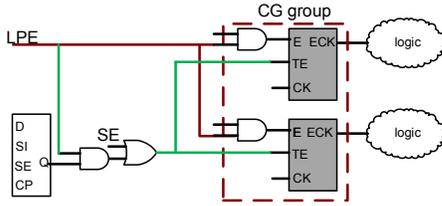


Figure 6. Extra control for CG group enable.

to not only set the corresponding LPE signal to 1, but also the functional logic driving the E pin of the CG, rendering it more difficult to figure out the proper values and often causes a drop in coverage. To mitigate this coverage loss due to the insertion of the power control logic, we create another CG group enable path to help ATPG tool to infer how to enable a CG group. It can be seen in Figure 6, that during capture, when LPE is set by the control logic, ATPG tool can enable the CG group by simply setting a "1" in the scan flop, other than setting the whole functional logic for the E pin. This will not violate the power budget because it is only enabled when LPE is set to 1.

Note that one may think it may be easier to directly get LPE signal ORed with SE without adding the scan flop. However, this will force the CG to be enabled by LPE, regardless the value from the functional logic on E pin, causing coverage loss on the functional logic.

3 Clock gate grouping

It can be seen from previous descriptions that the organization of CG groups is essential to the scheme, since CG group is the minimum set of flops that can be regulated for power control. This is done through a CG grouping method.

In this method, we first identify the CGs driving those scan flops that are hard to be exercised by ATPG. Since it is difficult for ATPG tool to toggle these flops, adding another level of constraint will cause excessive ATPG effort, more patterns and possibly lower coverage. Therefore, we leave these CGs as is and do not assign power control to them. Assume there are N CGs and let \mathbf{Seg} be the set of all CGs, i.e. $\mathbf{Seg}=\{CG_1, CG_2, \dots, CG_N\}$. We use $|CG_i|$ to

denote the number of flops controlled by CG_i . For each CG_i , we calculate an average probability P_{cg_i} , which represents how difficult the flops controlled by CG_i can be toggled by ATPG. A commercial ATPG tool is used to analyze CGs and simulate some random patterns to obtain a set of data, based on which an in-house tool is used to calculate P_{cg_i} . Details are ignored due to the lack of space. The CGs in \mathbf{Seg} are then sorted based on this probability such that $P_{cg_1} \leq P_{cg_2} \leq \dots \leq P_{cg_N}$.

We then choose a threshold value P_t , representing the percentage of total flops that will not be subject to our power control. This value is design specific, usually we can set it to the lowest granularity, e.g. 6% (or $\frac{1}{16}$). We will then remove the first K CGs from the sorted set \mathbf{Seg} . We select the largest K such that $\sum_{i=1}^K |CG_i| / \sum_{i=1}^N |CG_i| \leq P_t$. The CGs remained in \mathbf{Seg} are then randomly selected and placed into different CG groups. The sum of $|CG_i|$ in a CG group is counted until the power budget is reached. E.g., if we want the granularity of controllable power level to be 6.25%, we will need 16 CG groups. For each group, we will fill it with CGs from \mathbf{Seg} until 6.25% of total flops are included. We then start to fill the next CG group until all remaining CGs are placed into groups. Note that the flops eliminated from \mathbf{Seg} lead to a small "leaking" power, i.e. power not regulated by our controller. This will not affect the scheme significantly because such flops only constitute a small set and they are hard to be toggled by ATPG.

4 Experimental results

We insert the proposed transition power controller into several industrial designs and evaluate its effectiveness in pattern generation for transition faults using metrics such as capture power, overhead, test coverage etc. The controller uses three 4-to-16 decoders and one 2-to-4 decoder, hence it can provide a constraint of maximum capture power level at any combination of 6.25%, 6.25%, 6.25% and 25% of the full power. There are 16 CG groups, hence the finest granularity is 6.25%. Due to the lack of space, we ignore the details of the control logic truth table here and we only present results from three designs A, B and C in Table 3. All data are collected using a commercial ATPG tool and some in-house tools.

For each design, the first row lists the total number of scan flops and the number of controlled v.s. uncontrolled CGs. As shown in Section 3, most CGs will be controlled by our controller but a small portion of them will be left uncontrolled. The second row shows data from the original design without low capture power (LP) control. The next three rows show data after the proposed controller is inserted for three different power budgets, correspondingly. For each power budget, it is easy to infer the configuration of the controller through the set of the control bits. E.g. a

power budget of 31.25% can be obtained by enabling only one 2-to-4 decoder and one 4-to-16 decoder.

Column 2 presents the actual capture power, obtained by counting the average number of scan flops that are toggled by ATPG during capture, and then dividing this number by the total number of scan flops, hence it is a percentage. The actual number of toggling flops can be easily obtained from this value and the flop count. It can be seen that without control, the capture power is well over the feasible power budget, causing power droop. After controller is inserted, the actual power is constrained under the budget. An exception occurs in the second row of Design A. This is caused by the "leaking" power from the uncontrolled CGs. It is also because in this design, most of the flops under the enabled CGs can be easily toggled. Generally, a budget of 22.18% is still acceptable as compared with 18.75%.

Column 3 presents the transition faults coverage. We normalize the actual coverage value w.r.t. the original coverage without power control, which is therefore 100%. It can be seen that in all cases the coverage is not affected by the power constraints. In some cases it is even slightly higher because we inserted scan flops on CGs so that ATPG can enable the CGs more easily, as discussed in Figure 6.

Column 4 lists the pattern count. Intuitively, adding power constraints will cause larger pattern count. This is illustrated in Designs B and C, but the numbers are still quite manageable. In Design A, however, when the power budget is relieved to about 30%, the pattern count is almost identical to that of the original design. It indicates that we can reduce the peak capture power by almost 30% (from 44% to 31%) without hurting the pattern count, a great advantage.

In Column 5 it can be seen that the CPU time is several orders less than that using power-aware ATPG shown in Table 1. This renders the proposed scheme much more feasible for large industrial designs. Finally in Column 6 we give the estimated overhead of the controller, as a percentage of the entire design. The overhead is insignificant for all designs. As a conclusion, the proposed scheme effectively provides all features we list in Section 1.

5 Summary

In this paper, we have presented a novel technique for controlling the dynamic power dissipation during scan capture to reduce power droop in at-speed scan testing. We proposed the use of a transition controller with existing clock gates to limit the capture power within a preset power budget. We also presented a method to organize the clock gates into groups for controllability. Experimental results on industrial designs have shown that the proposed scheme can effectively constrain the capture power without significant increases in overhead and ATPG efforts.

Table 3. Experimental results.

	Power (%)	Coverage (%)	Pattern count	Time (sec)	Over-head(%)
Design A: 47579 flops, 1505 controlled/591 uncontrolled CGs					
No LP control	44.23	100	14232	4203	0.13
LP 18.75%	22.18	101	18255	3613	
LP 25%	24.28	101	16148	4220	
LP 31.25%	27.44	101	14496	3612	
Design B: 66076 flops, 1211 controlled/314 uncontrolled CGs					
No LP control	39.19	100	4206	603	0.09
LP 18.75%	17.8	100	9274	1202	
LP 25%	21.03	100	7617	602	
LP 31.25%	23.85	100	6210	602	
Design C: 92761 flops, 1994 controlled/621 uncontrolled CGs					
No LP control	36.37	100	6334	1807	0.065
LP 18.75%	16.41	100	20191	4204	
LP 25%	19.09	100	17001	2405	
LP 31.25%	22.7	100	13261	3016	

References

- [1] A. S. Abu-Issa and S. F. Quigley. Bit-Swapping LFSR and Scan-Chain Ordering: A Novel Technique for Peak and Average-Power Reduction in Scan-Based BIST. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 28, pp. 755C759, 2009.
- [2] Z. Abuhamdeh et al., Characterize Predicted vs Actual IR Drop in a Chip Using Scan Clocks. *Proc. Int. Test Conf.*, 21.1, 2006.
- [3] K. Chakravadhanula et al. Capture Power Reduction Using Clock Gating Aware Test Generation. *Proc. Int. Test Conf.*, 4.3, 2009.
- [4] D. Czysty et al. Low Power Scan Shift and Capture in the EDT Environment. *Proc. Int. Test Conf.*, 13.2, 2008.
- [5] V. R. Devanathan et al. PMScan : A Power-Managed Scan for Simultaneous Reduction of Dynamic and Leakage Power During Scan Test. *Proc. Int. Test Conf.*, 13.3, 2007.
- [6] B. Nadeau-Dostie et al. Power-Aware At-Speed Scan Test Methodology for Circuits with Synchronous Clocks. *Proc. Int. Test Conf.*, 9.3, 2008.
- [7] B. Li, L. Fang and M. S. Hsiao. Efficient Power Droop Aware Delay Fault Testing. *Proc. Int. Test Conf.*, 13.2, 2007.
- [8] C. Liu, V. Iyengar, and D. K. Pradhan. Thermal-Aware Testing of Network-on-Chip Using Multiple Clocking. *Proc. VLSI Test Symp.*, pp. pp. 46-51, 2006.
- [9] D. Mitra, et al., Test Pattern Generation for Power Supply Droop Faults. *Proc. Int. Conf. on VLSI Design*, pp. 343-348, 2006.
- [10] I. Polian et al. Power Droop Testing. *Proc. Int. Conf. Computer Design*, pp. 243-250, 2007.
- [11] C. P. Ravikumar, M. Hirech and X. Wen. Test Strategies for Low Power Devices. *Proc. Design, Automation and Test in Europe*, pp. 728-733, 2008.
- [12] L-T Wang, C. Stroud, N. Toubia etc.. System-on-Chip Test Architectures: Nanometer Design for Testability. , Morgan Kaufmann, San Francisco, Nov. 2007.

- [13] X. Wen et al. CTX: A Clock-Gating-Based Test Relaxation and X-Filling Scheme for Reducing Yield Loss Risk in At-Speed Scan Testing. *Proc. Asian Test Symp.*, pp. 397-402, 2008.