BUILT-IN PROACTIVE TUNING SYSTEM FOR CIRCUIT AGING AND PROCESS VARIATION RESILIENCE

A Thesis

by

NIMAY SHAH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2008

Major Subject: Computer Engineering

BUILT-IN PROACTIVE TUNING SYSTEM FOR CIRCUIT AGING AND PROCESS VARIATION RESILIENCE

A Thesis

by

NIMAY SHAH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Jiang Hu
Committee Members,	Weiping Shi
	Donald Friesen
Head of Department,	Costas Georghiades

August 2008

Major Subject: Computer Engineering

ABSTRACT

Built-In Proactive Tuning System for Circuit Aging and

Process Variation Resilience. (August 2008)

Nimay Shah, B.E., Dharmsinh Desai University

Chair of Advisory Committee: Dr. Jiang Hu

VLSI circuits in nanometer VLSI technology experience significant variations - intrinsic process variations and variations brought about by transistor degradation or aging. These are generally embodied by yield loss or performance degradation over operation time. Although the degradation can be compensated by the worst-case scenario based over-design approach, it induces remarkable power overhead which is undesirable in tightly power-constrained designs. Dynamic voltage scaling (DVS) is a more power-efficient approach. However, its coarse granularity implies difficulty in handling fine-grained variations. These factors have contributed to the growing interest in power-aware robust circuit design.

In this thesis, we propose a Built-In Proactive Tuning (BIPT) system, a low-power typical case design methodology based on dynamic prediction and prevention of possible circuit timing errors. BIPT makes use of the canary circuit to predict the variation induced performance degradation. The approach presented allows each circuit block to autonomously tune its performance according to its own degree of variation. The tuning is conducted offline, either at power on or periodically. A test pattern

generator is included to reduce the uncertainty of the aging prediction due to different input vectors.

The BIPT system is validated through SPICE simulations on benchmark circuits with consideration of process variations and NBTI, a static stress based PMOS aging effect. The experimental results indicate that to achieve the same variation resilience, proposed BIPT system leads to 33% power savings in case of process variations as compared to the over-design approach. In the case of aging resilience, the approach proposed in this thesis leads to 40% less power than the approach of over-design while 30% less power as compared to DVS with NBTI effect modeling.

DEDICATION

To all my loved ones

ACKNOWLEDGEMENTS

This thesis project has been a fitting crescendo to the wonderful two years I have spent at Texas A&M University. First and foremost, I would like to thank my advisor, Dr. Jiang Hu. Working with him has always been a thoroughly enriching and enthralling experience. It would have been impossible to complete this work without his unending support.

To my committee members, Dr. Weiping Shi and Dr. Donald Friesen, thank you for your invaluable suggestions and encouragement. Special Thanks to Dr. Ming Zhang at Intel Corporation for his feedback and suggestions.

To my good friend and guide, Rupak, whose insight and guidance have shaped this project.

To all the administrative staff members of the Department of Electrical and Computer Engineering, thank you for providing a comfortable and productive working environment.

To my classmates at some point of time or other, Charu, Shwetha, Karthik, Karan, Keerthi, Victor, Jampani, Ankit, thank you for making my life more interesting during homeworks, submissions and exams.

To the almighty, my friends, family and loved ones, as always, thank you for your selfless help, patience and love.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGEMENTS vi

TABLE OF CONTENTS vii

LIST OF FIGURES viii

1. INTRODUCTION 1

2. APPROACH 7

   2.1. BIPT: System Level Design 7
   2.2. Canary Circuit 8
   2.3. Test Pattern Generator and Control Circuit 11

3. EXPERIMENT SETUP AND RESULTS 19

   3.1. Common Experiment Setup Details 19
   3.2. Experiment Setup and Results for Process Variation Resilience 21
   3.3. Experiment Setup and Results for Aging Resilience 23

4. CONCLUSIONS 28

REFERENCES 29

VITA 33

LIST OF FIGURES

Page

Fig. 1. Classification of variations derived from [1] 2

Fig. 2. Overview of the proposed built-in proactive tuning system 7

Fig. 3. Canary circuit 9

Fig. 4. Error prediction and prevention using canary circuit 11

Fig. 5. Test pattern generator and control circuit 14

Fig. 6. Finish signal generator 15

Fig. 7. Gated clock circuit 15

Fig. 8. Operation of the clock stalling circuit shown in figure 7 16

Fig. 9. Generation of control signal to body bias block 18

Fig. 10. Power consumption for deterministic simulations 26

Fig. 11. Power consumption for statistical simulations considering the temporal variations of NBTI effect 26

LIST OF TABLES

Page

Table I Characteristics of ISCAS89 benchmarks under consideration: pre-BIPT processing and post-BIPT processing 20

Table II Power consumption for over-designed and BIPT schemes for s526 process variation simulations 24

Table III Power consumption for over-designed and BIPT schemes for s832 process variations simulations 25

1. INTRODUCTION

The achievement of device and interconnect parameter precision for present and future CMOS VLSI technologies is becoming an exponentially difficult task, thereby resulting in delay and power-consumption variability at the device, circuit and chip levels. Variability in device and process parameters will also continue to pose a challenge to continued scaling [1]. As CMOS VLSI technology in nanometer regime continues to scale aggressively for increased performance and integration density, designing robust low-power reliable systems in the presence of these variations becomes an increasingly daunting task [2].

Variations may be classified in several ways [1, 3, 4]. Figure 1 shows an elaborate classification based on [1] with excellent applicability to sub-nanometer technology nodes. The discussion on variations in this thesis is based on this classification. Variations may be spatial if they arise from the manufacturing process (sometimes also known as process variations) or temporal if they arise from device operation over time. Examples of spatial variations include random dopant fluctuation, sub-wavelength lithography induced variations [1, 4] etc. Random dopant fluctuations also impact the threshold voltage of the device [4]. Temporal variations may be classified further as reversible or irreversible. Environmental variations constitute reversible variations while variations brought about by transistor wear-out or aging mechanisms, such as NBTI (Negative Bias Temperature Instability) and HCI (Hot Carrier

---

This thesis follows the style and format of *IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems.*

Injection) [1, 4, 5] are a part of irreversible variations. NBTI manifests itself by degradation of PMOS threshold voltage [6, 7] whereas HCI results in threshold voltage increase in mostly NMOS transistors. When technology scales from 180nm to 65nm and beyond, the MTTF (Mean Time To Failure) of processors due to aging effects is reduced by about 76% [6]. That is, if a chip would have previously lasted for 10 years, now it can perform well for about 2 years. Process variations have resulted in an increasingly lower yield if not taken care at design stage by means such as design for manufacturability (DFM) [8]. Therefore, it becomes increasingly imperative to address these issues in chip design.

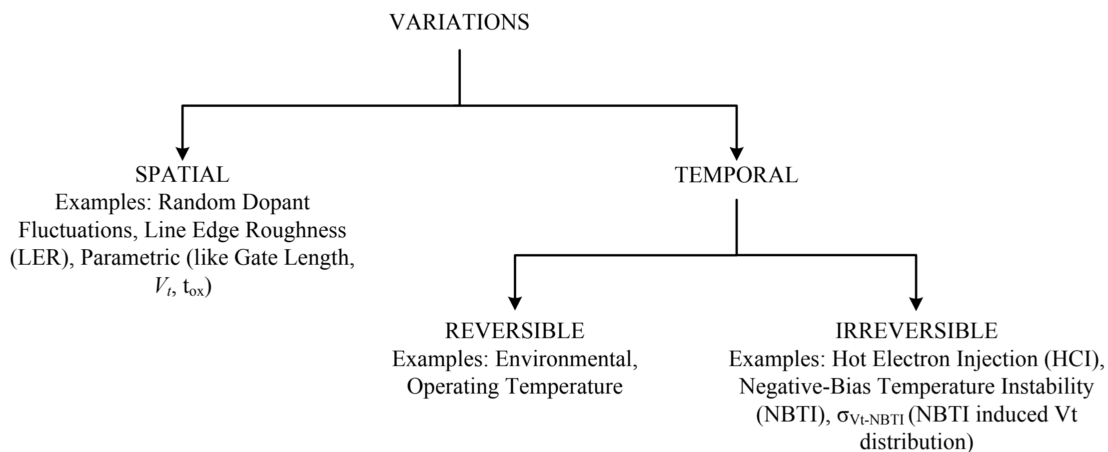

Fig. 1. Classification of variations derived from [1]

To handle the process variation and circuit degradation problem in Deep Sub-Micron (DSM) chips, designers generally resort to corner analysis and design the circuit such that it is guaranteed to perform in the ‘worst-case’. Two common approaches are to use a conservative supply voltage or to over-size transistors such that the aging or

process variation degraded performance can still meet specifications [7]. These approaches are able to provide a guard band against projected process variations or extend chip lifetime under the aging effect. However, they inevitably increase circuit power dissipation and therefore the designer hits another wall of nanometer integrated circuits – the increasingly tight power constraint. Conservative voltage supplies are set for the worst-case scenarios which seldom occur and thus mean unnecessarily large currents and power dissipation in the more frequently occurring typical case scenarios. Over-sized transistors usually imply unnecessarily large timing slack and therefore wasteful power dissipation during for the circuits with negligible performance degradation. Alternatively, architectural approaches [5, 6] are suggested for mitigating the variation problem. One technique is architectural-level adaptation [5, 6] such as DVS (Dynamic Voltage Scaling) which is an improvement over setting a permanently high supply voltage. For instance, a chip can operate at relatively low supply voltage level when new and switch to higher supply voltage level when it gets aged or detects process variation induced errors. Such adaptation can avoid the wasteful power. However, this is a coarse-grained technique – the supply voltage level is usually fixed for major partitions of the chip, if not across the entire chip. In general, the variations and their effects vary among different components of a circuit. In order to ensure the performance of an entire chip, the DVS must be performed according to the worst transistor aging or the worst process variation margin. That is, even though only 1% transistors may be strongly affected due to variations, the chip-level supply voltage has to be increased although the other 99% transistors have suffered very minor variations.

This thesis proposes a Built-In Proactive Tuning (BIPT) system to mitigate the effects of these issues in a power-efficient manner. This system includes a canary circuit which can generate predictive warning signals for performance degradation. According to the warning signal, circuit speed is tuned through body bias such that the performance degradation is compensated. The proactive tuning is performed offline, at power on or periodically. Since aging is a slow change with a time constant of weeks/months and process variations are a static change, periodic tuning of once in a few days is sufficient to capture these changes. The offline tuning has the advantage of allowing relatively easy control on input vectors. When detecting performance degradation or circuit delay variation, one must consider the delay uncertainty due to different input vectors. Even if there is no warning signal for certain input vectors, there is still risk of delay errors under other input vectors. Therefore, a Test Pattern Generator (TPG) is used in the system in order to have large input vector coverage. TPG is usually a part of Built-in Self Test (BIST) hardware; thus, if a chip already has BIST circuit, TPG does not cause extra overhead. The proposed BIPT system has the following advantages:

- It can be applied at circuit block level instead of the chip level architectural approach [5, 6]. In other words, each block can be tuned according to its own degree of variation. Evidently, the finer granularity control allows improved power efficiency.
- Its performance degradation detection is obtained from the actual operating circuit as opposed to replica circuit in other adaptive design methods [9]. Since

the detection is more direct, it is more reliable. Using TPG further improves the reliability of the detection.

- Its proactive nature can avoid the complex error correction schemes in retroactive systems [10, 11]. The retroactive systems rely on pipeline flush [11] or instruction replay [10] and therefore are restricted to processor designs. In contrast, our system can be applied to both processors and general sequential circuits.

The BIPT scheme proposed in this research is described in section 2 and has the advantages as listed above. To the best of our knowledge, the existing approaches fail to capture all of these advantages into a single solution. The work of [9] is a block level adaptive body bias technique. However, its delay variation detection is obtained from replica circuits which often have discrepancy from the actual operating circuits. The Razor based techniques [3, 4] use direct variation detection, but they rely on complex error correction method and are restricted to processor designs. Another retroactive method [12] is mainly targeted for fast variations such as voltage variations and hence complements our work. The canary circuit based predictive detection is proposed in [8]. However, it is applied with online tuning which suffers from delay uncertainty due to different input vectors. The recent work of [13] focuses on only the aging detection instead of an overall tuning system. Actually, the detection method in [13] can be easily adopted in our tuning system.

The remainder of the thesis is organized as follows. Section 2 explains the Built-In Proactive Tuning System in detail. The system level details are described first. Finer

details of operation are explained next, starting off with canary circuit operation [8] and moving on to circuit level details for the Control Circuit and the TPG. In Section 3 the experiment setup is described and the experimental results are explained. Finally this thesis is concluded in Section 4.

2. APPROACH

This section describes the approach presented in this thesis for process variation and circuit-aging resilient design: Built-In Proactive Tuning (BIPT) System. This section is divided into three sub-sections. Section 2.1 describes the BIPT system at an architectural level. Section 2.2 explains the error prediction mechanism in canary circuit and section 2.3 describes the circuit level details of the TPG and the control circuit.

2.1. BIPT: SYSTEM LEVEL DESIGN

The Built-In Proactive Tuning (BIPT) System consists of the existing main circuit augmented with a Test Pattern Generator (TPG), Body Bias Circuitry, Canary Circuit and Control circuit. Figure 2 shows these blocks and the corresponding interface signals.

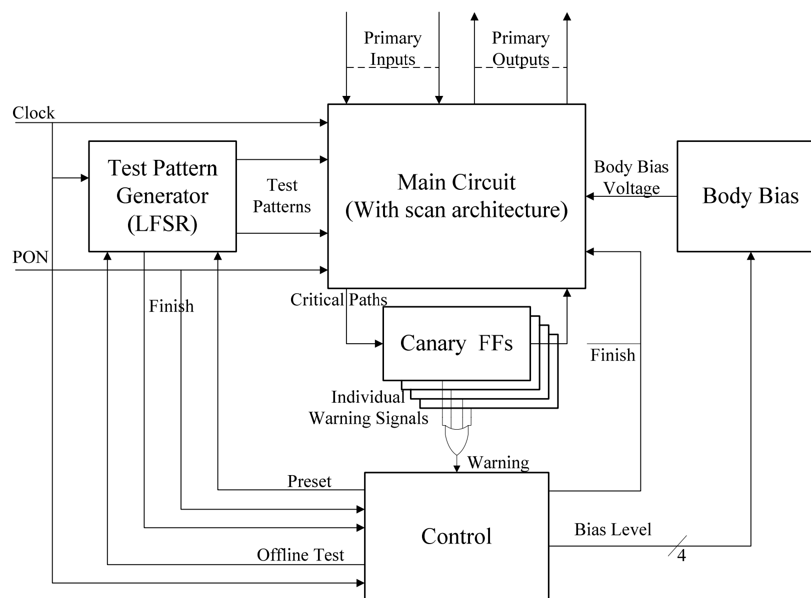

Fig. 2. Overview of the proposed built-in proactive tuning system

At power-on or periodically, the BIPT system can launch test vectors from TPG and then tune circuit body voltage according to the observations from the canary circuit. As described in section 2.2, canary circuit plays the role of detecting performance degradations. A *Warning* signal is generated by the canary circuit when the timing constraint is tight on one or more of the few critical paths where these are inserted. The top-level warning signal is the *OR* of all the individual canary circuit warning signals.

The Linear Feedback Shift Register (LFSR) [14] is implemented as a pseudo-random test pattern generator which applies these random patterns when offline test is in progress. It is triggered by the *preset* signal from the control block.

The control block monitors the status of all the blocks and issues control signals. *PON* is the power-on-reset signal which is an active high reset signal issued on start-up and basically triggers the offline test. *Offline test* is an active high signal indicating that offline test is in progress. The most critical activity performed by the control block is to monitor the warning signal from the canary circuit. Whenever the warning signal goes high, the control block appropriately sets the body bias to selective gates on the critical paths circuit. This function is performed by the bias level signal passed by the control block to the body bias block. This interface and the body bias block are modeled as in [9]. The body bias is adaptive to the circuit state: it automatically selects from four available options of forward body bias using a counter – decoder based scheme.

2.2. CANARY CIRCUIT

The canary circuit [8] is for detecting variation-induced performance degradation in a predictive manner. As shown in Figure 3, a canary circuit consists of two flip-flops;

a main FF and a canary FF. The main FF gets the direct input and the canary FF which serves as the checker part gets the input through a delay buffer. This delay in the input reaching the two flops serves as the guard band for error detection. The outputs from these flops are fed to a *XOR* gate which functions as a comparator, outputting 1 when these are different and thereby predicting the occurrence of an error. Some advanced designs of canary circuits are proposed in [4, 13].

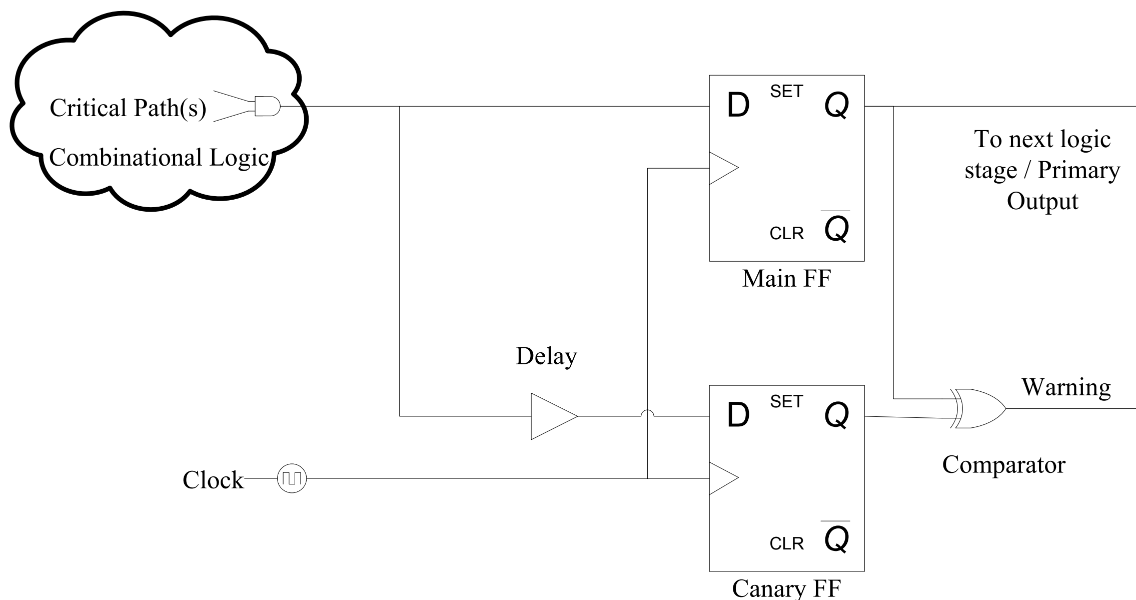

Fig. 3. Canary circuit

Canary circuit is a typical case design alternative of Razor [11]. However, in contrast to Razor, which delivers a delayed system clock to the checker part (shadow FF), canary circuit delivers a delayed input signal to the checker part (canary FF). This simplifies the clock tree synthesis and routing as there is just one system clock now. Also, the delay buffer placed before the canary flip-flop always has a positive delay,

even if affected by process variation or aging, which makes the canary flip-flop recover from variation induced effects by itself. Canary circuit also predicts timing errors rather than detecting them afterwards. The predictive warning allows the user to take preventive measures before the timing violation actually occurs and thus the system does not run into any corrupt data states, except for errors that cannot be predicted such as single event upset (SEU) errors. However, timing violations caused by the variations of our interest can be predicted effectively by architectures such as canary circuit [13].

The waveform in figure 4 shows how the approach of this thesis uses the canary circuit to predict and prevent the occurrence of timing violations. Following the labels used in figure 3, *Data* is the input to the *Main FF* and *Delayed Data* is the input to the *Canary FF*. On the first clock edge, the main FF clocks in Data1 and the canary FF clocks in Data0 thereby causing a mismatch between their outputs and raising the Warning signal high. This indicates that the canary circuit has predicted a timing error. As described in the overview of the BIPT system, this forward biases selective transistor(s) on the critical path(s). As a result Data1 is sped up through the combinational path and it arrives at the Main FF such that the timing requirements are satisfied at the second clock edge. Thus, the error is prevented from occurring. In the absence of any error prevention mechanism, the data at the main FF would've been as shown in red color in figure 4. The authors in [15] explain how Razor would handle this erroneous situation.

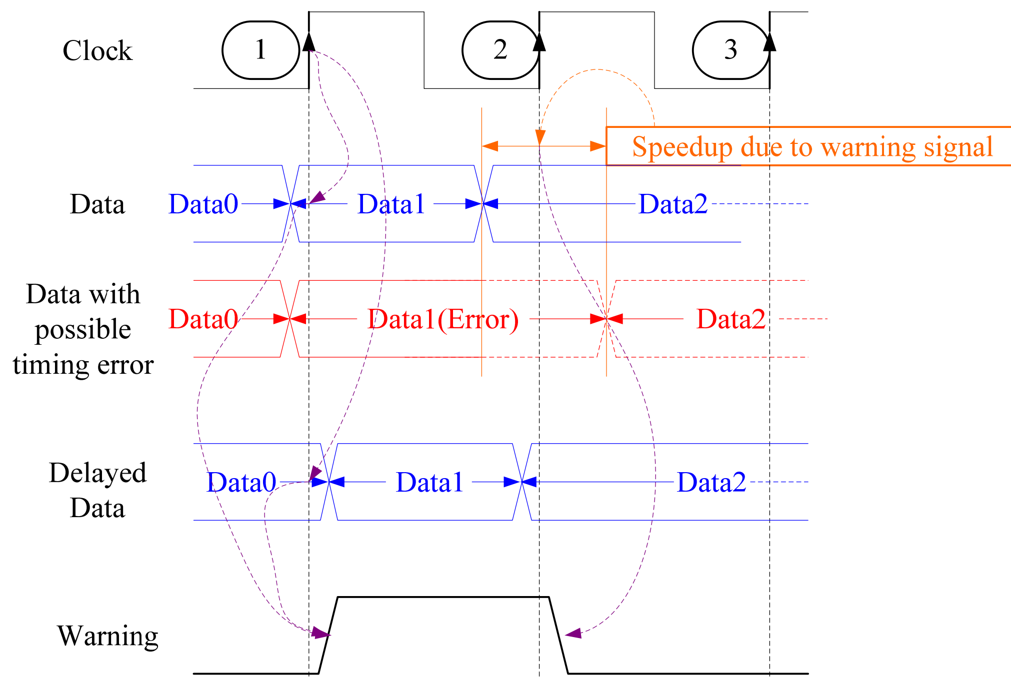

Fig. 4. Error prediction and prevention using canary circuit

2.3. TEST PATTERN GENERATOR AND CONTROL CIRCUIT

Figure 5 shows the gate level implementation of the control circuit. *Finish* signal going high indicates the completion of offline testing, *PON* is the power-on-reset signal, *Warning* is the timing error prediction signal from the canary flip-flop and *Preset* is the active low signal to set the flip-flops in the LFSR to high state on power-on-reset. The preset generation circuit is shown in the dotted box in figure 5. The initial states of all the flip-flops in the LFSR on power-on-reset is '1', thus the starting seed for the LFSR is all 1's. The LFSR shown in figure 3 is a 12-bit LFSR; it implements a primitive polynomial to generate 4095 patterns (2^n-1; n=12) before returning back to the initial

state of all 1's. The outputs of the flip-flops in the LFSR are fed to a scan chain through a mux-d connection. These connections are omitted in figure 5 for clarity.

The control circuit is triggered by the power-on-reset signal (PON), which remains high for one cycle on each power on of the chip. On each power-on-reset, the *offline test signal* triggers the offline testing. Generation of the offline test signal is shown in a box in figure 5. The *offline test signal* is the input for preset generation circuit that presets the flip-flops in the LFSR to high, which serves as the initial seed for the test patterns generated by the LFSR. The *Finish* signal is generated by the circuit shown in figure 6. Its first stage consists of a 12-input *AND* gate and the second stage consists of a 2-input Muller-C gate. Muller-C gate is an AND gate for events i.e., it produces a high output when all the inputs are high and goes low when all the inputs transit to low state. The description about Muller-C gates can be found in [16]. As shown in figure 6, the outputs of the flip-flops in the LFSR are connected to a 12-input *AND* gate. The output of this *AND* gate and $\overline{PON}$ feed to a 2-input Muller-C gate to produce the finish signal. On every power-on-reset, the flip-flops of the LFSR are preset to 1, thus the output of the *AND* gate rises high. Since PON is active high, $\overline{PON}$ is low at startup and thus finish stays at 0 initially. $\overline{PON}$ stays high for one clock cycle and then goes to low. When all the 4095 test patterns have been generated, the output of the *AND* gate goes high again and since $\overline{PON}$ is also high; finish goes high indicating the completion of offline testing. After finish goes to high, at the next clock edge, the output of the *AND* gate goes low due to a pattern other than all ones. However, the finish signal still stays high because of the property of the Muller-C gate to hold the previous value

until both the inputs transition to the same value. In this case, although the output of the *AND* gate goes low, since $\overline{PON}$ is still high, finish stays high.

The possible timing violations in the critical paths, i.e., the paths that are affected due to aging and process variation, are predicted by the warning signal from the canary circuit. To prevent these timing violations from occurring, the data through these critical paths has to be sped up. This requires the applications of suitable forward body bias voltage. Since, the body bias generation circuit takes some time to apply correct bias to the devices on these critical paths, the LFSR needs to be stalled. In the approach described, the clock to the LFSR is stalled by using gated clock circuitry shown in figure 7. In figure 7, the circuit can stall the clock for one clock cycle, which is sufficient to change the body bias of the devices on the critical paths. However, if more time is needed then the clock can be stalled for a longer period of time using cascaded Muller-C gates.

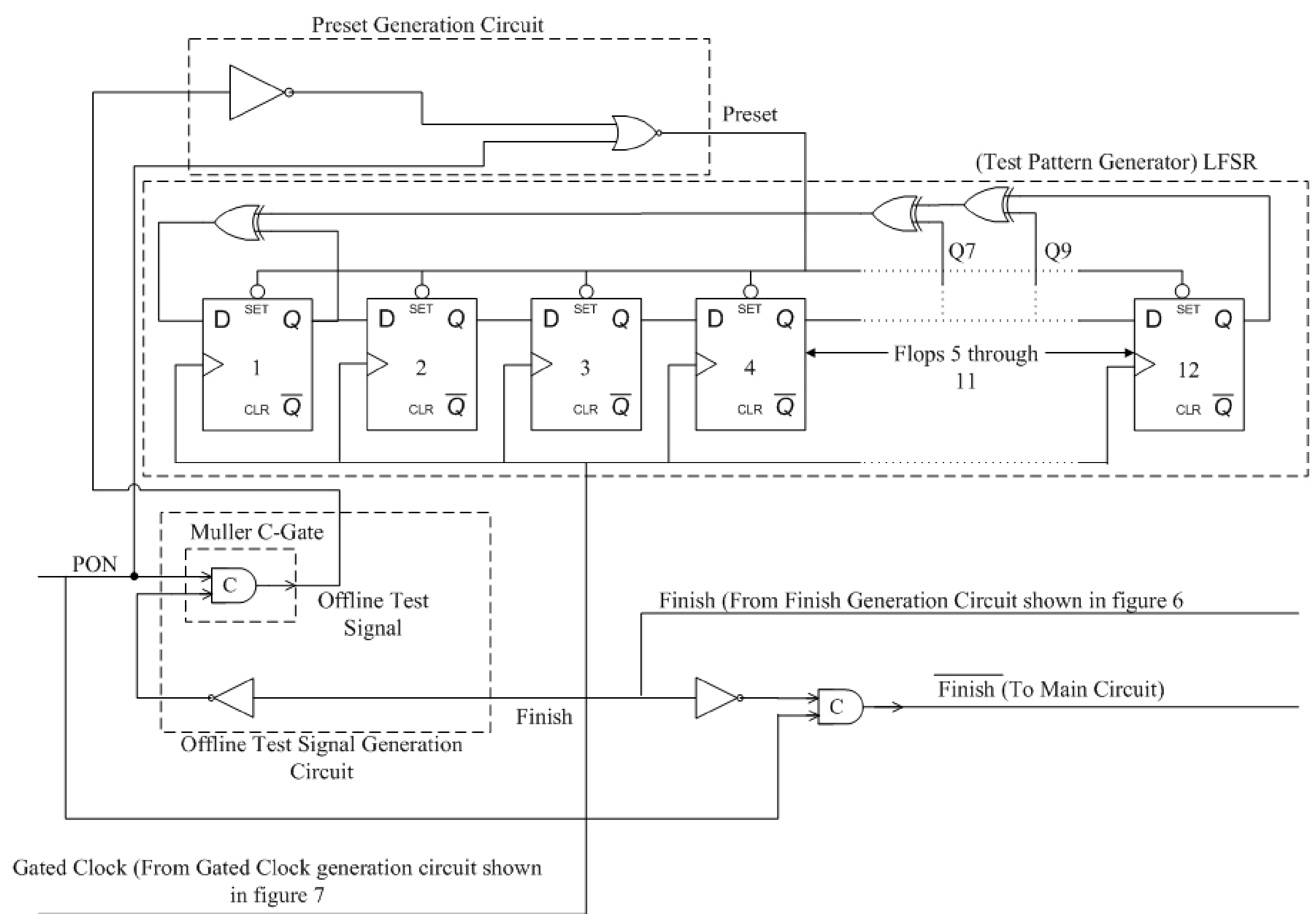

Fig. 5. Test pattern generator and control circuit

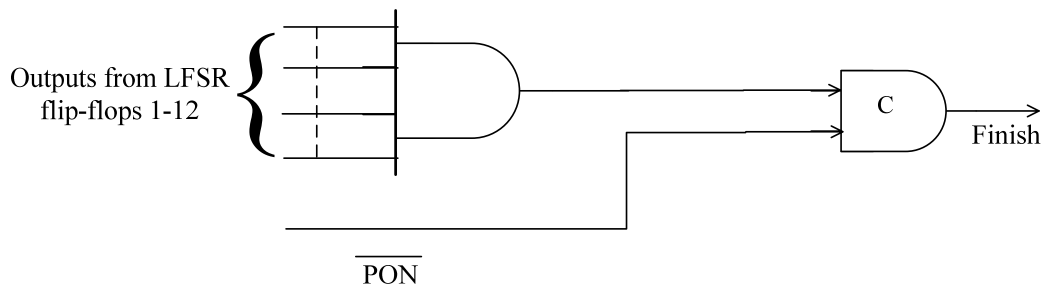

Fig. 6. Finish signal generator

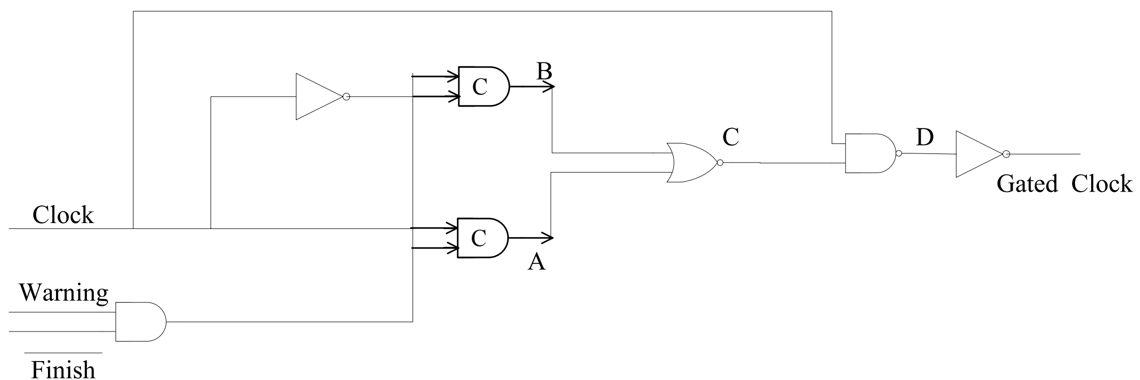

Fig. 7. Gated clock circuit

Figure 8 describes the operation of the circuit in figure 7. The input signals, *Clock* and *Warning*, are as shown. *Finish* is 0 throughout offline testing and has been omitted over here. The signals at different points of the clock-gating circuit are also shown. *A* is the output of the Muller-C gate with inputs as $\overline{Clock}$ and warning. Similarly, *B* is the output of the Muller-C gate with inputs as *Clock* and warning. *A* and *B* are *NOR-ed* to obtain *C* which is *AND-ed* with clock to get the final output signal, which is the *Gated Clock*. The gated clock has the desired characteristic of stalling the

clock for a period of time derived from the time taken for the change of body bias to the transistors on the critical path(s).

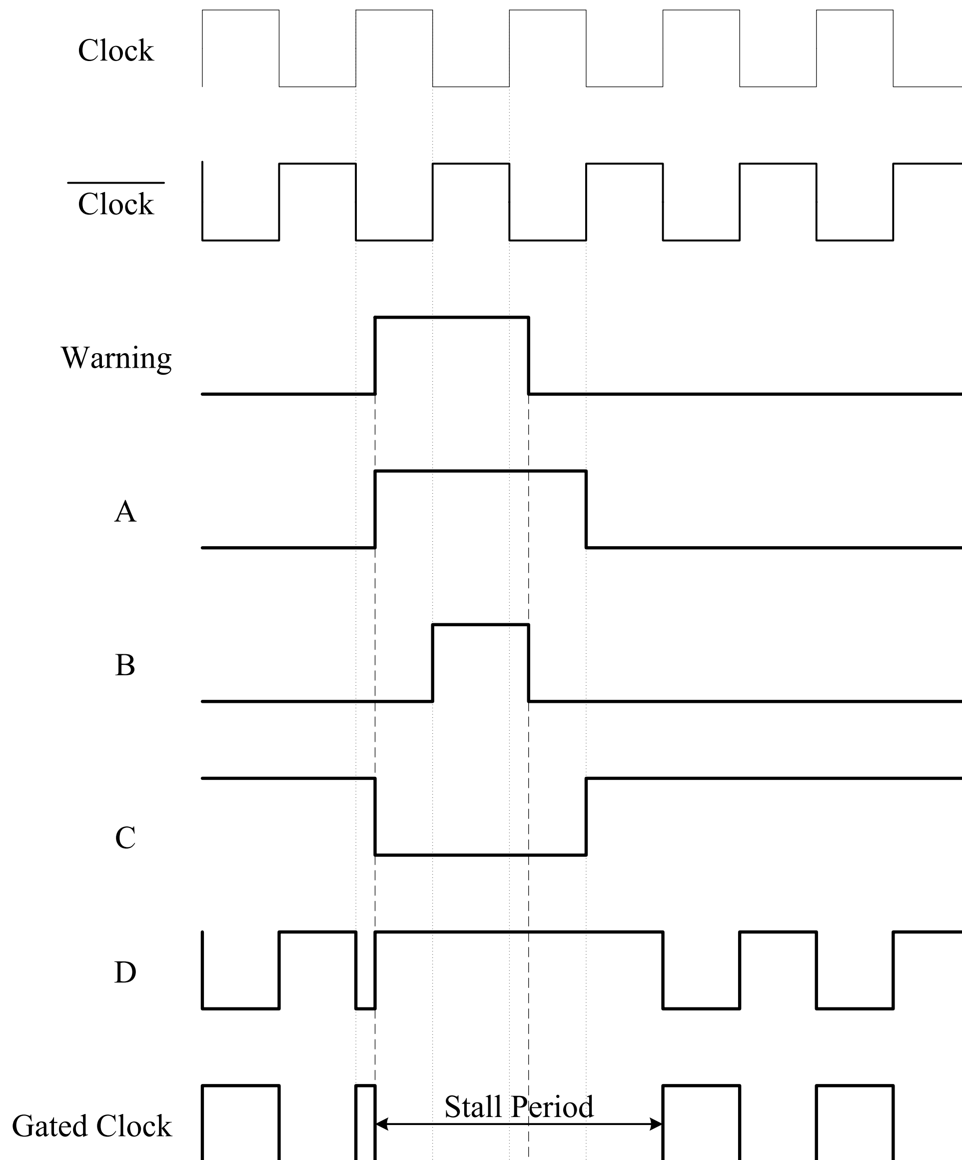

Fig. 8. Operation of the clock stalling circuit shown in figure 7

Figure 9 shows the body-bias generation logic. Two flip-flops are used in the body bias generation circuit, thus the body bias can be chosen from one of the four possible body bias levels. Level 0 is the no-bias condition; levels 1 to 3 are in the increasing order of the forward body biases. Since intrinsic process variations are more or less variations that remain constant throughout the lifetime of the chip, a single body-bias level is sufficient to correct them. However, when dealing with aging degradation, which monotonically degrades the circuit performance with time, forward body bias or reduced reverse body bias is necessary to restore the circuit performance. To be able to handle both cases efficiently, an up counter is used that counts upward (increases forward bias) when a warning signal is generated by the canary circuit. It counts upward till it reaches the highest forward body bias state (binary 11 in our case) and freezes in that state. A four-state counter is implemented as few forward body bias levels are sufficient for the circuits under consideration. However, larger number of forward bias levels can be generated by adding extra flip-flops in the body bias generation circuit. The outputs Q1 and Q2 of the counter go to a 2-to-4 decoder. The decoder outputs are inputs to the body bias circuitry which is implemented as in [9] and enable the appropriate body bias option.

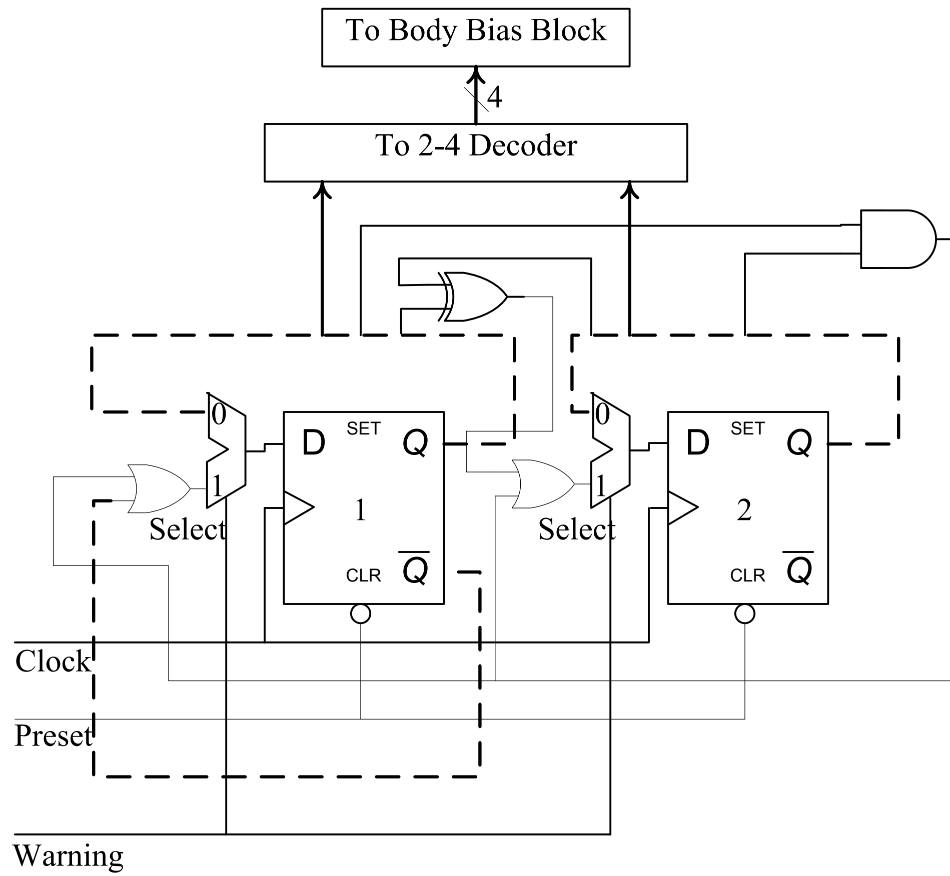

Fig. 9. Generation of control signal to body bias block

3. EXPERIMENT SETUP AND RESULTS

3.1. COMMON EXPERIMENT SETUP DETAILS

The experiments for offline testing are performed on ISCAS89 sequential benchmarks: s526 and s832 [17]. First, we augment these circuits with BIPT hardware. To do this, we determine the critical paths in these circuits by using a static timing analyzer written in C. The gate libraries needed for the static timer are characterized in HSPICE for 90nm model card from BPTM [18]. The flip-flops at the output of the critical paths are replaced by canary circuits (consisting of a main FF and a canary FF), whose structure and operation is described in section 2.2. Once the placement of canary circuits in the netlist is determined, the paths from input of the canary circuit are traversed in a breadth first fashion till we reach either a flip-flop or a primary input and insert the body-bias contacts for the gates on these paths. The flip-flops are replaced by mux-d scan flops and extra scan flip-flops are added for the primary inputs. The scan flip-flop has two inputs, one input is connected to the input of the original flip-flop and the other input is connected to the output of the LFSR. A scan-enable signal is used to select between the two inputs. $\overline{Finish}$ signal serves as the scan-enable for scan flip-flops in our design. It can as well be a user-defined input. The characteristics of the benchmarks pre- and post-BIPT processing are shown in Table 1. Column 3 shows the number of flip-flops originally in the design and column 7 shows the number of these flip-flops replaced by canary circuits respectively. Column 6 shows the number of mux-d scan flops inserted in the design.

The other important task is to set the clock period for simulations and thus set the target performance for both the benchmarks. The clock period for the simulations is determined by applying a pre-defined V_{dd} to just the main circuit (without BIPT hardware). For this V_{dd}, we run simulations to find out the nominal clock period such that no error occurs during offline testing. We add a safety margin of 15% to this period and the resulting clock period becomes clock period for our simulations. For a V_{dd} of 1.15V, this final value is found to be 480ps (f = 2.08 GHz) and 600ps (f = 1.67 GHz) for s526 and s832, respectively. Since a circuit with BIPT hardware doesn't need to operate with high safety margins, the clock period is set to be 480ps (600ps) and for this clock period the minimum V_{dd} is determined such that no error occurs during offline testing. For both the benchmarks, V_{dd} is set to 0.925v for BIPT.

Table I
Characteristics of ISCAS89 benchmarks under consideration: pre-BIPT processing and post-BIPT processing

ISCAS89 Benchmark	No of Gates	No of flip-flops (FF) (pre-BIPT processing)	No of Primary Inputs	No of Primary Outputs	No of mux-d scan-flops (post-BIPT processing)	No of FF replaced by canary circuits (post-BIPT processing)
s526	193	21	3	6	12	4
s832	262	5	18	19	12	2

One other important task is to set the value of the delay element inside the Canary circuit. This serves as in-situ safety margin for the BIPT system. Through

simulations this was determined to be equal to 50ps for both the benchmarks. This is a reasonable number as it is in the 10-15% range of the clock period used for simulations. Since this is not such a big delay, it is simply implemented by a 2-inverter chain with a sizing ratio of four. By a sizing ratio of four we mean that if the first inverter is taken to be unit-sized the second inverter is four times this size.

To validate the BIPT system on the benchmarks, two sets of experiments are performed. For the first set of experiments, threshold voltage (V_t) variations and gate length variations arising from intrinsic process variations are considered. For the other set of experiments, the effect of NBTI induced PMOS V_t degradation in these circuits is considered. Simulations take into account the effect of both nominal V_t degradation and temporal variations in V_t degradation; using models as described in [19]. These simulations are carried out in HSPICE [20] at a simulation temperature of 100° *C*.

3.2. EXPERIMENT SETUP AND RESULTS FOR PROCESS VARIATION RESILIENCE

Simulations for process variation resilience are performed by drawing samples using *Latin Hypercube Sampling* (LHS) [21, 22], which is a fast Monte Carlo technique [21]. Monte Carlo analysis more often than not requires a large number of random sampling points. This results in expensive overall simulation cost especially if the simulation time is large [21], which is true in our case. LHS utilizes the cumulative distribution function of the random variable x to select the random sampling points in a controlled manner. Thus, instead of selecting random samples from a random number generator like Monte Carlo, LHS ensures that the sampling points are distributed all over

the random space, ensuring better estimation accuracy [21, 23]. A MATLAB subroutine is used to generate the desired LHS samples [24]. Full correlation is assumed between transistors of the same gate and zero correlation is assumed between transistors of different gates. The variations are assumed to follow Gaussian distribution such that the ±3σ limits are chosen to be equal to 15% of the nominal value. Twenty sets of samples are prepared, where each sample consists of *n* Latin-Hypercube samples, *n* being the number of gates in the benchmark under consideration. The nominal values of PMOS threshold voltage (V_{tp}), NMOS threshold voltage (V_{tn}) and gate length are taken as -$0.303V$, $0.2607V$ and 90nm, respectively.

The efficiency of the BIPT scheme is demonstrated by comparing it with the over-design case as the baseline case. As the name suggests, the over-design case does not have any error prediction or error recovery mechanism and is thus designed with a safety margin to ensure the circuit operates error-free in the worst-case corner. The power consumed for both cases is observed and similar set of simulations are carried out for both the benchmarks. The power consumption of BIPT system includes power dissipation due to the TPG, canary circuit and control circuit. The results are as tabulated in Tables 2 and 3. Table 2 reports the power consumption for s526 and Table 3 shows the power consumption for s832. For both the tables, Column 1 shows which LHS generated spice deck is under consideration. Column 2 reports the power consumption in *mw* for the over-designed case while Column 3 reports the power consumption using the BIPT scheme. Column 4 reports the power savings by using BIPT scheme over the over-

design technique in percentage. The last row in both the tables reports the averages. On an average, the BIPT scheme consumes 33% less power than the over-design case.

3.3. EXPERIMENT SETUP AND RESULTS FOR AGING RESILIENCE

The simulation for device aging variation is broken up into two parts: (a) Deterministic Simulations and (b) Statistical Simulations. Deterministic simulations are carried out for 0%, 5% and 10% of NBTI induced deterministic V_t degradation. We compare power consumed by BIPT scheme with the over-designed case as the baseline case. The power estimation of BIPT system includes power dissipation due to the TPG, canary circuit and control circuit. The over-design implemented here is a conservative scaling of V_{dd} level. In particular, the V_{dd} for the over-designed case is set such that it does not cause timing violations and meets the performance targets at 10% V_t degradation as well. This value is found to be 1.2V for both s526 and s832 for 2.08 GHz and 1.67 GHz respectively. On the other hand, BIPT scheme allows for typical case circuit design, and adapts to the degradation of the circuit during its lifetime. Thus, the operating voltage is kept at 0.925V for BIPT simulations. Figure 10 plots the power consumed for deterministic simulations for s526 and s832. From the simulation results, we can observe that, on an average, BIPT scheme leads to power savings of 45% compared to the over-designed case.

Table II
Power Consumption for over-designed and BIPT schemes for s526 process variation simulations

Latin Hypercube Sample Set Number	Power Consumption (mW)		% Power Saving
	Over-Designed Scheme (A)	BIPT Scheme (B)	
1	24.97	16.01	35.88
2	24.62	16.10	34.61
3	25.04	16.14	35.54
4	24.55	16.01	34.79
5	24.50	16.01	34.65
6	25.32	16.09	36.45
7	25.04	16.03	35.98
8	25.35	16.10	36.49
9	24.55	16.04	34.66
10	25.44	16.12	36.64
11	24.94	16.03	35.73
12	24.52	16.03	34.62
13	25.37	16.03	36.82
14	25.58	16.02	37.37
15	24.72	16.11	34.83
16	25.21	16.01	36.49
17	25.03	16.05	35.88
18	24.86	16.01	35.60
19	24.48	16.01	34.60
20	24.49	16.04	34.50
Average	24.93	16.05	35.61

Table III
Power consumption for over-designed and BIPT schemes for s832 process variations simulations

Latin Hypercube Sample Set Number	Power Consumption (mW)		% Power Saving
	Over-Designed Scheme (A)	BIPT Scheme (B)	
1	21.23	14.58	31.32
2	21.24	14.52	31.64
3	21.27	14.59	31.41
4	21.26	14.60	31.33
5	21.24	14.60	31.26
6	21.28	14.57	31.53
7	21.27	14.59	31.41
8	21.22	14.56	31.39
9	21.25	14.55	31.53
10	21.25	14.54	31.58
11	21.25	14.58	31.39
12	21.31	14.60	31.49
13	21.26	14.53	31.66
14	21.27	14.52	31.73
15	21.27	14.58	31.45
16	21.28	14.59	31.44
17	21.21	14.54	31.45
18	21.31	14.59	31.53
19	21.25	14.57	31.44
20	21.26	14.59	31.37
Average	21.26	14.57	31.47

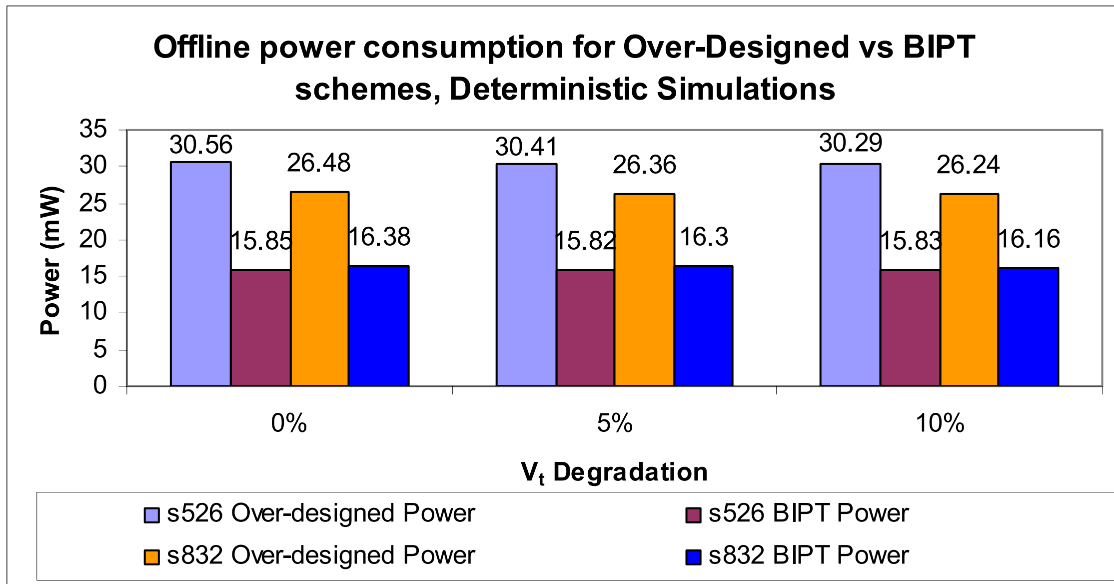

Fig. 10. Power consumption for deterministic simulations

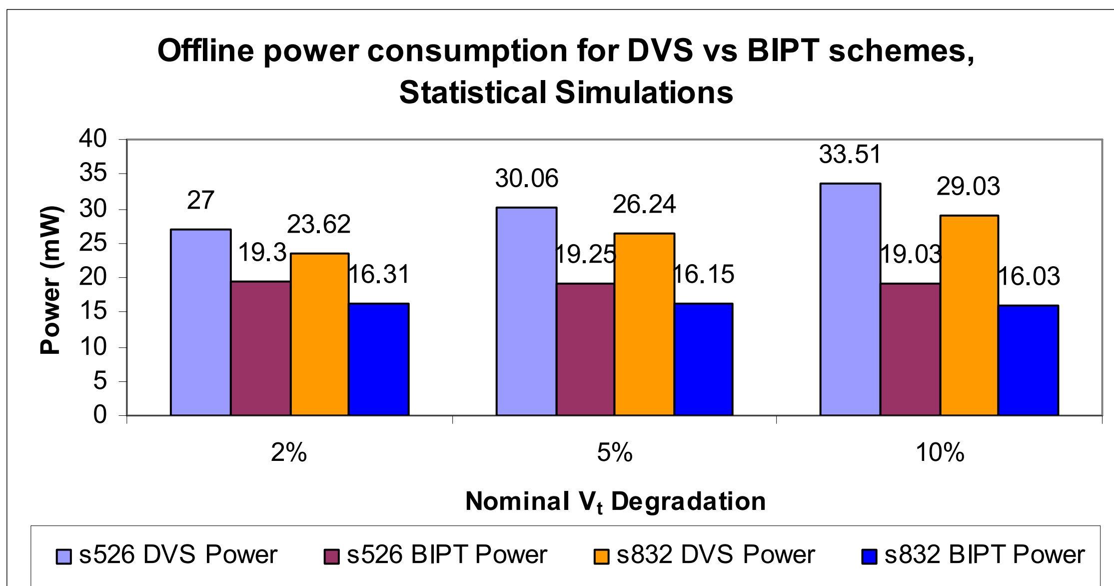

Fig. 11. Power consumption for statistical simulations considering the temporal variations of NBTI effect

For statistical simulations, the temporal variation in lifetime V_t degradation is accounted for. The lifetime V_t degradation is modeled as a Poisson random variable, which takes into account the statistical variation in the underlying process causing V_t degradation [5]. Power consumed by implementing BIPT scheme is compared with the Dynamic Voltage Scaling (DVS) scheme. Thus, the dynamic voltage scheme serves as the baseline for statistical simulations. The simulations are carried out for statistical V_t variation over 2%, 5% and 10% of nominal value. The V_{dd} values for DVS are selected such that in each nominal case, the circuit is ensured to work for the worst statistical variation. Thus, for a transistor whose V_t is degraded by 5% (temporal variation) over and above the 2% nominal degradation, V_{dd} is selected such that the circuit would still work without any timing violations if all transistors in the circuit were similarly affected. The operating voltages for the DVS schemes are found to be 1.15V, 1.2V and 1.25V for 2%, 5% and 10% degradations respectively. The operating voltage for BIPT case still remains at 0.925V. Figure 11 plots the power consumed for statistical simulations for s526 and s832. From the experimental results, it can be observed that, on an average, BIPT scheme leads to power savings of 30% compared to the dynamic voltage scaling approach. The average power saving here is less than the previous case because Dynamic Voltage Scaling scheme is an improvement over the over-designed approach. The power for DVS methodology increases as V_t degradation increases because of the fact that the voltage supply is varied keeping in mind the most degraded transistor.

4. CONCLUSIONS

In this thesis, a novel typical-case power-aware, robust and reliable design technique, Built-In Proactive Tuning (BIPT) system, is presented. Built-In Proactive Tuning system allows VLSI circuits to autonomously compensate for process variation and aging-induced performance degradations. Being a typical-case design methodology and the ability to tune itself, it helps the designer to avoid the unnecessary safety margins in the design stage. As a result, BIPT consumes 33% less power than the over-design methodology when considering process variations. Due to its adaptive nature, BIPT is power-efficient and uses about 45% less power than over-design based aging compensation. Since it is a middle-grained approach, it can achieve 30% power reduction compared to the coarse-grained DVS method. Thus, the proposed design technique has excellent applicability in the current era as low power and reliable system design becomes increasing challenging with the rapid technology scaling in VLSI circuits.

REFERENCES

[1] K. Bernstein, D. J. Frank, A. E. Gattiker, W. Haensch, B. L. Ji et al., "High-performance CMOS variability in the 65-nm regime and beyond," *IBM Journal of Research and Development,* vol. 50, no. 4, pp. 433-450, 2006.

[2] A. Agarwal, "Process variation aware high performance low power VLSI system design in nano-scale regime," Purdue University, Lafayette, IN, 2005.

[3] E. Malavasi, S. Zanella, J. Uschersohn, M. A. M. M. Misheloff, and C. A. G. C. Guardiani, "Impact analysis of process variability on digital circuits with performance limited yield," in *6th IEEE International Workshop on Statistical Methodology*, 2001, pp. 60-63.

[4] M. Zhang, T. M. Mak, J. Tschanz, K. S. Kim, N. Seifert et al., "Design for resilience to soft errors and variations," in *Proceedings of the 13th IEEE International On-Line Testing Symposium*, 2007, pp. 23-28.

[5] J. Srinivasan, S. V. Adve, P. Bose, and J. A. Rivers, "The case for lifetime reliability-aware microprocessors," in *Proceedings of the 31st Annual International Symposium on Computer Architecture*, 2004, pp. 276-287.

[6] J. Srinivasan, S. V. Adve, P. Bose, and J. A. Rivers, "Lifetime reliability: toward an architectural solution," *IEEE Micro,* vol. 25, no. 3, pp. 70-80, 2005.

[7] B. C. Paul, K. Kang, H. Kufluoglu, M. A. Alam, and K. Roy, "Negative bias temperature instability: estimation and design for improved reliability of nanoscale circuits," *IEEE Transactions on Computer-Aided Design of Intergrated Circuits and Systems,* vol. 26, no. 4, pp. 743-751, 2007.

[8] T. Sato, and Y. Kunitake, “A simple flip-flop circuit for typical-case designs for DFM,” in *Proceedings of the 8th International Symposium on Quality Electronic Design*, 2007, pp. 539-544.

[9] J. W. Tschanz, J. T. Kao, S. G. Narendra, R. Nair, D. A. Antoniadis et al., “Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage,” *IEEE Journal of Solid-State Circuits,* vol. 37, no. 11, 2002.

[10] K. A. Bowman, J. W. Tschanz, N. S. Kim, J. C. Lee, C. B. Wilkerson et al., “Energy-efficient and metastability-immune timing-error detection and instruction-replay-based recovery circuits for dynamic-variation tolerance,” in *IEEE International Conference on Solid-State Circuits*, San Francisco, 2008, pp. 402-403.

[11] S. Das, D. Roberts, S. Lee, S. Pant, D. Blaauw et al., “A self-tuning DVS processor using delay-error detection and correction,” *IEEE Journal of Solid-State Circuits,* vol. 41, no. 4, pp. 792-804, 2006.

[12] R. Samanta, G. Venkataraman, N. Shah, and J. Hu, “Elastic timing scheme for energy-efficient and robust performance,” in *9th International Symposium on Quality Electronic Design*, 2008, pp. 537-542.

[13] M. Agarwal, B. C. Paul, M. Zhang, and S. Mitra, “Circuit failure prediction and its application to transistor aging,” in *25th IEEE VLSI Test Symposium*, 2007, pp. 277-286.

[14] M. Abramovici, M. Breuer, and A. Friedman, *Digital Systems Testing and Testable Design*, New Jersey: IEEE Press, 1990.

[15] D. Ernst, K. Flautner, T. Mudge, N. S. Kim, S. Das et al., “Razor: a low-power pipeline based on circuit-level timing speculation,” in *Proceedings of the 36th Annual IEEE/ACM International Symposium on Microarchitecture*, 2003, pp. 7-18.

[16] I. E. Sutherland, “Micropipelines,” *Communications of the ACM,* vol. 32, no. 6, pp. 720-738, 1989.

[17] F. B. D. Bryan, and K. Kozminski, “Combinational profiles of sequential benchmark circuits,” in *Proceedings of the International Symposium on Circuits and Systems*, 1989, pp. 1929-1934.

[18] Y. Cao, T. Sato, M. Orshansky, D. Sylvester, and C. Hu, “New paradigm of predictive MOSFET and interconnect modeling for early circuit simulation,” in *Proc. CICC*, 2000, pp. 201-204.

[19] K. Kang, S. P. Park, K. Roy, and M. A. Alam, “Estimation of statistical variation in temporal NBTI degradation and its impact on lifetime circuit performance,” in *Proceedings of the 2007 IEEE/ACM international conference on Computer-aided design*, 2007, pp. 730-734.

[20] L. Nagel, “Spice: A computer program to simulate computer circuits,” *University of California, Berkeley UCB/ERL Memo M*, pp. 201-204, 1995.

[21] X. Li, J. Le, and T. P. Lawrence, *Statistical performance modeling and optimization*, Hanover, MA, USA: Now Publishers, 2006.

[22] M. Stein, "Large sample properties of simulations using latin hypercube sampling," *Technometrics,* vol. 29, no. 2, pp. 143-151, 1987.

[23] J. P. Edzer, and B. M. H. Gerard, "Latin hypercube sampling of Gaussian random fields," *Technometrics,* vol. 41, no. 4, pp. 303-312, 1999.

[24] B. Minasny, "Latin Hypercube Sampling ", *MATLAB Central File Exchange*, 2004.

VITA

Nimay Shah received his Bachelor of Engineering degree in electronics and communication from Dharmsinh Desai University, in India in 2006. He graduated with his Master of Science degree in computer engineering from the Department of Electrical and Computer Engineering at Texas A&M University in August 2008. During his graduate studies he has done research in various aspects of VLSI circuit design including robust circuit design, typical case circuit design, variation resilient design, and low-power circuit design.

Nimay Shah may be reached at the Department of Electrical and Computer Engineering, 315D WERC, Texas A&M University, College Station, TX 77843. His email address is: nimay_shah@tamu.edu.