

Improved State Integrity of Flip-Flops for Voltage Scaled Retention under PVT Variation

Sheng Yang, Saqib Khursheed, Bashir M. Al-Hashimi, *Fellow, IEEE*, David Flynn, Geoff V. Merrett

Abstract—Through measurements from 82 test chips, each with a state retention block of 8192 flip-flops, implemented using 65-nm design library, we demonstrate that state integrity of a flip-flop is sensitive to process, voltage, and temperature (PVT) variation. It has been found at 25°C that First Failure Voltage (FFV) of flip-flops varies from die to die, ranging from 245-mV to 315-mV, with 79% of total dies exhibiting single bit failure at FFV, while the rest show multi-bit failure. In terms of temperature variation, it has been found that FFV increases by up to 30-mV with increase in temperature from 25°C to 79°C, demonstrating its sensitivity to temperature variation. This work proposes a PVT-aware state-protection technique to ensure state integrity of flip-flops, while achieving maximum leakage savings. The proposed technique consists of characterization algorithm to determine minimum state retention voltage (MRV) of each die, and employs horizontal and vertical parity for error detection and single bit error correction. In case of error detection, it dynamically adjusts MRV per die to avoid subsequent errors. Silicon results show that at characterized MRV, flip-flop state integrity is preserved, while achieving up to 17.6% reduction in retention voltage across 82-dies.

Index Terms—State Integrity, State Retention, Voltage Scaling, Online Error Detection and Correction, Leakage Power Reduction.

I. INTRODUCTION AND RELATED WORK

SUPPLY voltage scaling is an effective technique for reducing standby mode leakage power, and is frequently used in energy-constrained designs [1]. Recent research has shown that it effectively reduces sub-threshold and gate-leakage power in deep-submicron designs [1], [2]. This is because of negative exponential relationship of leakage power and supply voltage, when $V_{gs} \approx 0$ [3].

Energy constrained designs that require low wake-up time (from Standby¹ to Active State) can retain their flip-flop values during sleep state to reduce wake-up time. Power minimization through supply voltage scaling during sleep state was proposed in [4], [5], where voltage scaling is implemented by using IR-drop of diode to reduce power supply. Minimum Retention Voltage (MRV) of a flip-flop is defined as a voltage value, such that supply voltage scaling below MRV leads to vulnerable flip-flop state. For a given flip-flop design, minimum retention

voltage is characterized across all process and temperature corners. Using 130-nm technology node, MRV of a flip-flop has been studied, and ‘canary’ flip-flop is proposed to provide a safety margin for voltage scaling, as it fails earlier than core flip-flop; it can be used to trade-off state-integrity with leakage power savings [6]–[8]. Other than flip-flop, it was recently shown through Monte-Carlo SPICE simulations that process variation also has an impact on minimum retention voltage of storage elements such as cache and SRAM [9]–[11]. Post silicon characterization method was proposed for fine-grain cache line supply voltage control [9] and for SRAM retention voltage adjustment in individual die [10], [11].

This paper presents measured results on 82 dies to demonstrate flip-flop state integrity challenges due to process, voltage and temperature (PVT) variation. It uses silicon results to propose a PVT-aware state-protection technique, which consists of the following two parts: firstly, a binary search based MRV (Minimum Retention Voltage) characterization algorithm is proposed, and it is used to determine MRV of individual die in the presence of PVT variation. Secondly, a control flow is proposed for state monitoring and protection of flip-flops, which uses parity for multi-bit error detection and single bit error correction. Silicon results show that state integrity is preserved, while reducing leakage power during standby mode.

This paper is organized as follows: test chip implementation is discussed in Sec. II. Test chip measurement results to demonstrate that state integrity of a flip-flop is sensitive to process, voltage, and temperature (PVT) variations in sleep state are shown in Sec. III. Proposed PVT-aware state-protection technique and related silicon results to demonstrate leakage power saving with state retention integrity are presented in Sec. IV. Finally the paper is concluded in Sec. V.

II. TEST CHIP

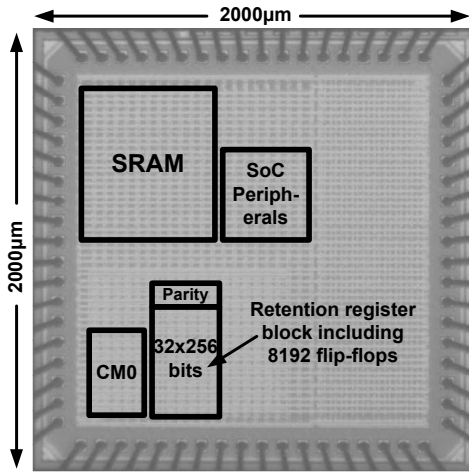
To analyze state-integrity of voltage scaled state retention flip-flop, a register array of 8192 flip-flops referred as retention register block is implemented in TSMC 65nm “LP” low leakage technology with nominal operating voltage of 1.2V using Unified Power Format (UPF) design flow and standard EDA tools (Synopsys, Mentor Graphics). The test chip is shown in Fig. 1, where Fig. 1-(a) shows the die photo of the test chip, and Fig. 1-(b) shows the test board photo. Measurement presented in this work are based on 82 test chips. As shown in Fig. 1-(a), the retention register block is located on bottom left-hand side of the die, and the parity storage is placed above the register block. ARM CortexTM-M0 micro-controller [12] referred as CM0 is used in this

Manuscript received on 24th Sept, 2012, revised on 18th Jan, 2013. This work is supported by Engineering and Physical Sciences Research Council (EPSRC), UK, in part under grant nos. EP/H011420/1 and EP/G067740/1.

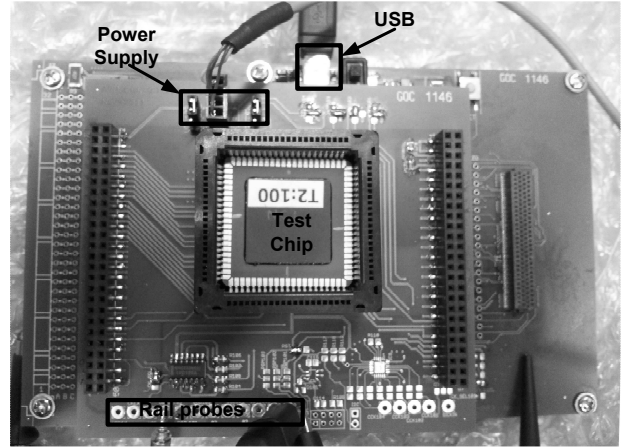
S. Yang, S. Khursheed, B. M. Al-Hashimi and G. V. Merrett are with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ, UK. (email: sy08r@ecs.soton.ac.uk; ssk@ecs.soton.ac.uk; bmah@ecs.soton.ac.uk; gvm@ecs.soton.ac.uk)

D. Flynn is with ARM Limited, Cambridge, UK. (email: david.flynn@arm.com)

¹Onwards, Standby Mode is referred as “Sleep State”.



(a) Test silicon



(b) Test board

Fig. 1. Test silicon fabricated and packaged for evaluation.

work for state monitoring and it can be seen on the left-hand side of register block. A pair of oscillators have also been implemented to measure delay variation due to process, voltage and temperature variations, when considering inter-die and intra-die process variation. The oscillators are located next to the parity storage unit. Due to its small size, it is not marked on the die photo. The implementation is part of a 2x2-mm system on chip (SoC), the rest of the SoC is made up of SRAM for instruction and data storage. The test board is shown in Fig. 1-(b), which provides rail probes, power supply connections and USB interface to communicate with the host computer through an ASCII debug protocol.

III. STATE INTEGRITY CHALLENGES

Fig. 2 shows the schematic of master-slave flip-flop commonly used in modern digital designs. The master latch is transparent when clock is low and the slave latch is transparent when clock is high. The slave latch that is made of two cross coupled inverters is used for state retention at low supply voltage. In theory, the latch is capable of retaining its state at a very low supply voltage, given the design is not effected by process variation, that is both PMOS and NMOS of the two cross coupled inverters (I2 and EI2) have the same drive strength; and there is no noise for example due to, supply

voltage fluctuation, radiations, substrate and inductive noise. This is because transistor ON-current is always higher than its OFF-current.

In this work, First Failure Voltage (FFV) of a flip-flop is defined, such that the supply voltage at FFV leads to the first bit(s) failure in a design consisting of n flip-flops, where bit-failure refers to the change in stored logic value from the initial (or correct) value of flip-flop. Note, single or multiple bits failure is possible at FFV. Due to process, voltage and temperature variation, FFV of a given design varies from die-to-die. To ensure state integrity, it is important to analyze this change in FFV of state-retention flip-flop. Using measured results from 82 dies, this section analyzes the change in FFV due to process, voltage and temperature variation. Section III-A shows FFV distribution from 82 dies. Section III-B analyzes change in FFV due to within die process and voltage variation, and finally Section III-C analyzes change in FFV due to temperature variation.

A. Effect of Process Variation across Dies

Using measured results from 82 dies, Fig. 3 shows the spread of First Failure Voltage (FFV). This measurement was carried out at room temperature (25°C) using 82 dies, each with 8192 flip-flops, and with the implementation setup shown in Fig. 1. For this measurement, test board was connected with a host computer through USB interface, and Python script was used as a control software to communicate between the host computer and the test board. The FFV is found using a binary search algorithm with resolution of 1-mV per iteration, starting from 400-mV, until first bit failure is observed (Fig. 11). Each iteration consisted of the following five steps: 1) Voltage of the design was set to 1.2-V, 2) A single logic value (logic-0 or logic-1) was stored in all 8192 flip-flops, referred as *initial logic state*, 3) Supply voltage was reduced to a lower voltage with a fall time of 40-µs and this was held for 10-sec, 4) Supply voltage was raised back to 1.2-V with a rise time of 40-µs, 5) Flip-flop states were observed and compared with the *initial logic state* to determine if FFV has been observed.

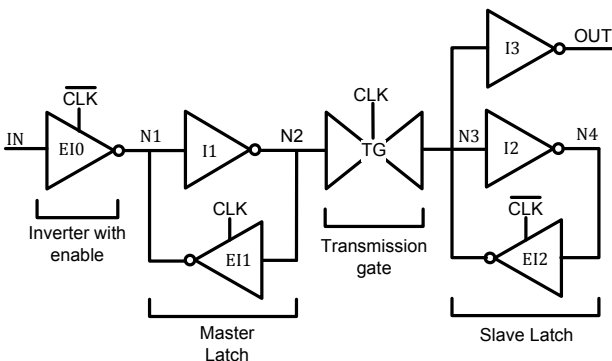


Fig. 2. Schematic of Master-Slave Flip-Flop.

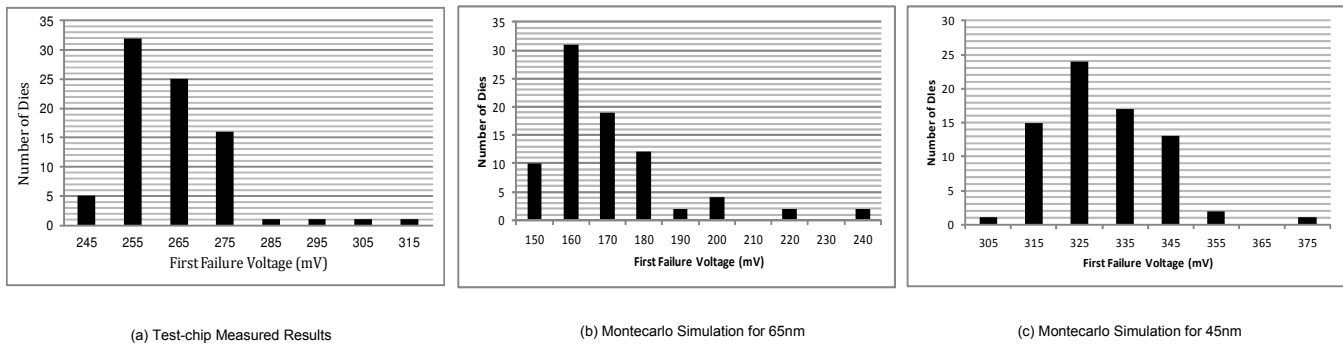


Fig. 3. Results from 82 Dies showing the spread of First Failure Voltage Point of voltage-scaled flip-flop for state retention in room temperature (25°C).

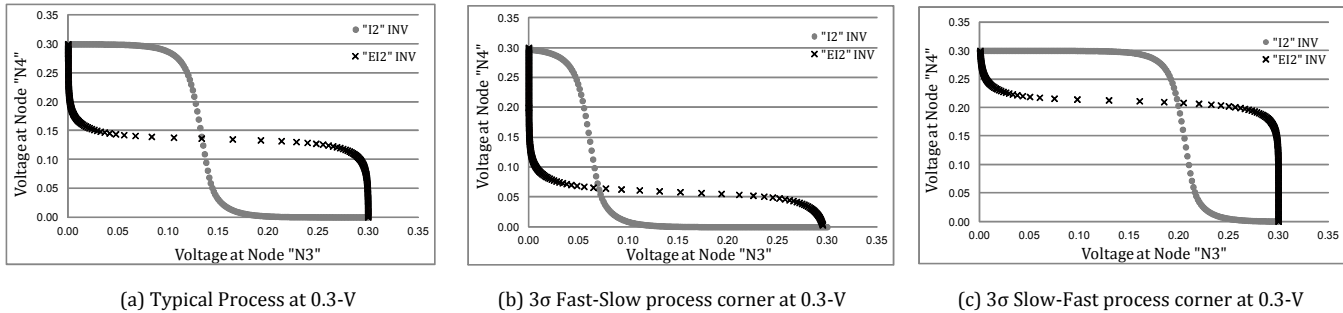


Fig. 4. Simulated results showing Noise Margins of a typical flip-flop for state retention (Fig. 2), when operating at 0.3-V, and reduced Noise Margins due to process variation at Fast-Slow and Slow-Fast corners.

Each iteration was executed 10-times to avoid the effect of jitter and most common value was recorded. The maximum jitter of 2-mV was observed when charge time was 40- μ s and this value increased to 18-mV with a shorter charging time (≤ 4 - μ s). From Fig. 3, it can be seen that the First Failure Voltage (FFV) point of each design varies from die-to-die, and for these 82 dies FFV is in between 245-mV to 315-mV, with 95% dies exhibiting their FFV below 285-mV. The measured results shown in this work use Standard Vth (SVT) cell. Simulation results using Low Vth (LVT) and High Vth (HVT) cells show that LVT cell has the lowest FFV and HVT cell has the highest FFV. Fig. 3-(b) shows the Monte-Carlo simulation result for the same 65-nm TSMC technology

library. This is used to analyze how well simulated results correlate with measured results (Fig. 3-(a)). The effect of process variation is incorporated by varying three parameters, which include: gate length (L), threshold voltage (V_{th}), and mobility (μ_{eff}) (Mobility varies due to variation in effective strain in a strained silicon process [13]). These parameters follow Gaussian distribution ($\pm 3\sigma$ variation) with standard deviations of 4% for L, 5% for V_{th} and 21% for μ_{eff} . It can be seen that the overall distribution trend remains the same while the mean FFV has shifted to lower voltage. This is because simulation results do not take into account environmental noise and inductive effects. It can also be observed that the spread of FFV is slightly wider in simulation than measured results, this

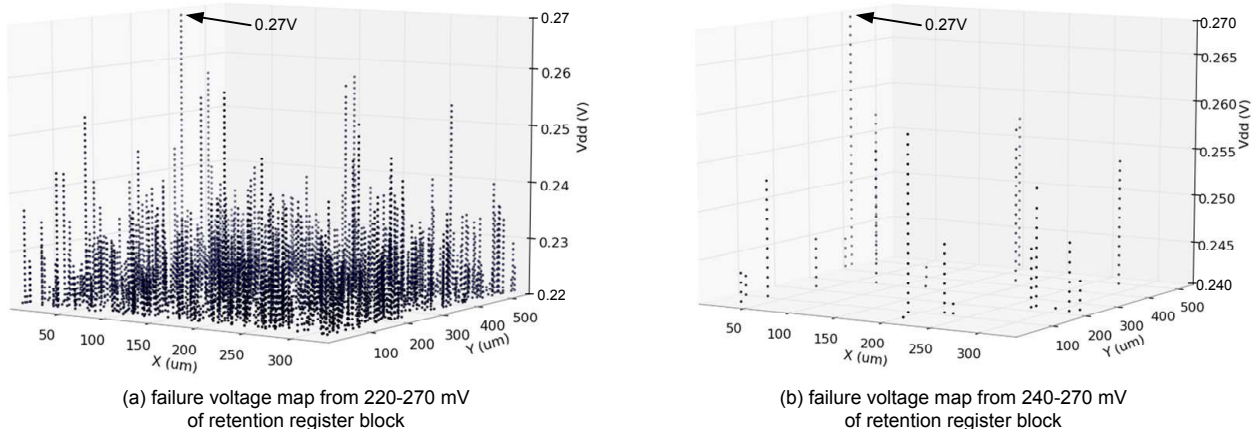


Fig. 5. Measured results showing failure bit locations in the retention register block of the test chip.

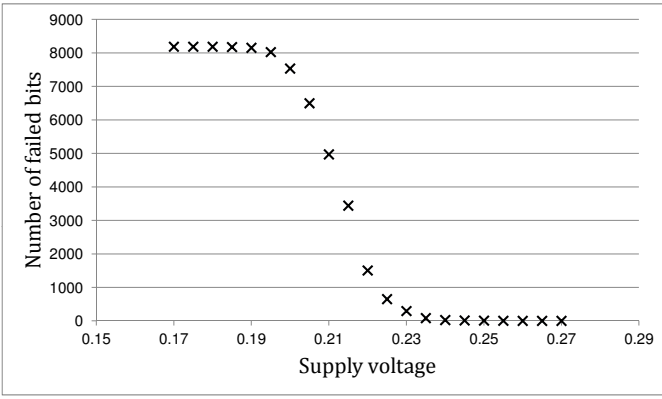


Fig. 6. Measured results showing the distribution of failing voltage point of flip-flops at reduced supply voltage.

is because the effect of process variation on fabricated devices is less than simulated results. Similarly, we also simulated FFV for a 45-nm technology library [14]. Results are shown in Fig. 3-(c). When comparing it with simulated results of 65-nm technology library (Fig. 3-(b)), it can be observed that the overall distribution trend remains the same, however the mean FFV has shifted to a higher voltage due to higher process variation.

To get an insight into FFV spread of state-retention flip-flops. We simulated the effect of process variation on state-retention capability of a flip-flop. The simulation is carried out on the slave latch (Fig. 2), using a design from typical, fast-slow and slow-fast process corners of TSMC 65-nm Low-Power design library. The results are shown in Fig. 4. In all three plots, the voltage transfer curve of inverter ‘I2’ is represented by circular dots, and that of ‘EI2’ is represented by black crosses. X-axes show voltage at node ‘N3’ and y-axes show voltage at node ‘N4’ (output, Fig. 2). Fig. 4-(a) shows a typical design operating at 0.3-V, it can be seen that the logic threshold voltages of both inverters are equal at about $0.5 * V_{dd}$, leading to symmetric noise margins for storing both logic values in the slave latch. However, when considering a fast-slow process corner, operating at 0.3-V (Fig. 4-(b)), the logic threshold voltages of both inverters reduce to about $0.23 * V_{dd}$, leading to asymmetric noise margins for storing logic-0 on both nodes ‘N3’ and ‘N4’. This means that a small noise can convert logic-0 to logic-1 on ‘N3’ and ‘N4’, leading to data corruption of stored states at low-supply voltage. Similarly, when considering a slow-fast process corner, operating at 0.3-V (Fig. 4-(c)), the logic threshold voltages of both inverters increase to about $0.7 * V_{dd}$, again leading to asymmetric noise margins for storing logic-1 on both nodes ‘N3’ and ‘N4’ and a small noise can convert logic-1 to logic-0 on both nodes (‘N3’ and ‘N4’), leading to state corruption at low-supply voltage.

These results clearly demonstrate that the state-retention capability of a voltage scaled flip-flop is affected by process variation, and simulated results (Fig. 4) reveal that due to process variation, noise margin of a flip-flop gets skewed leading to variation in FFV (First Failure Voltage) as observed in measured results from 82 dies shown in Fig. 3.

B. Effect of within Die Process and Voltage Variation

To analyze failure voltage across 8192 flip-flops within a single die, we used a die exhibiting nominal process characteristic and measurement setup outlined in Sec. III-A. Fig. 5 shows measured results from the test chip to demonstrate the failure voltage behaviour of 8192 flip-flops and their individual XY co-ordinates within the design layout (Fig. 1-(a)). Fig. 5 shows the location of failed flip-flops as observed on the test chip. ‘X’ and ‘Y’ axes show physical location of each flip-flop and indicates the distance (in μm) from the bottom left corner of the retention register block (Fig. 1). Z-axis shows the supply voltage during retention mode. It can be observed that the First Failure Voltage (FFV) occurs at 270-mV, and this flip-flop continues to fail with further reduction in supply voltage. A few subsequent failure points are at about 260-mV. In general, over all flip-flops, when the supply voltage is ≤ 240 -mV, more flip-flops start to fail, and this is shown in Fig. 6 using 5-mV step size. For this measurement (Fig. 6), ten test runs were conducted, and the plot shows the average number of failed bits over all test runs. Fig. 6 shows that the first bit failure is observed at 270-mV, and the number of failed bits increase with further reduction in supply voltage until the supply voltage is reduced to 190-mV, where all flip-flops failed to retain initially stored logic values.

To get an insight into failure pattern across all 82 dies, a measurement is taken at room temperature (25°C) to determine the FFV of each die, and voltage difference between the first and subsequent failing flip-flops. The results are shown in Fig. 7, where X-axis show FFV of each die, and Y-axis show the voltage difference between first and subsequent failing flip-flop for each die. For example, in case of Die-3, FFV is observed at about 250-mV, and the voltage difference (Y-axis) is 0, representing multi-bit failure (two or more flip-flops) at FFV. Similarly, in case of Die-2, FFV is observed at about 270-mV, but the difference between the first bit failure and subsequent bit failure is about 9-mV. As can be seen, the voltage difference between first and subsequent flip-flop failures is highest (66-mV) in case of Die-1. *In general, across all dies it was found that 20.73% of all dies show multiple bit failure at First Failure Voltage (FFV) point and the rest (79.27%) show only single bit failure at FFV. This finding is exploited in our proposed technique (Sec. IV), which employs a simple parity based error detection and correction technique for multi-bit error detection and single bit error correction.*

C. Effect of Temperature Variation

The effect of temperature variation on state retention voltage of a flip-flop was also examined from three dies marked in Fig. 7. These three dies represent both nominal and corner cases. We measured first failure voltage (FFV) on the following four temperatures: 25°C , 41°C , 56°C , and 79°C . The temperature of the test chip was raised using a temperature chamber. Fig. 8 shows the relationship between FFV and temperature. For all three dies, as expected, it was found that FFV increases with temperature. This is because transistor leakage current increases with temperature, while drive current (I_{on}) decreases [3]. In case of the master-slave flip-flop shown

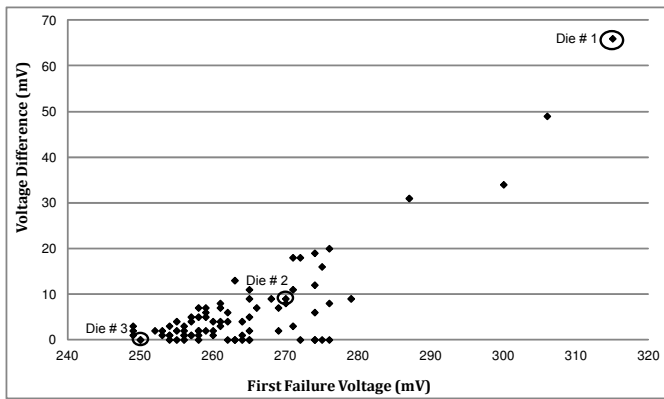


Fig. 7. Measured Results from 82 dies showing Voltage Difference between First Flip-Flop Failure Voltage and Subsequent Flip-Flops' Failure Voltages.

in Fig. 2, the state integrity of a storage node (N3 or N4) depends on the charge stored and the feedback current. As temperature increases, this feedback current reduces due to increase in leakage current and reduction in drive current, which negatively affects state retention capability of storage node at higher temperatures. This means that the state retention voltage of a flip-flop has to be raised at higher temperatures to ensure state-integrity.

To get an insight into the combined effect of process, voltage, and temperature variation on a given design. Using Die-2 (Fig. 7), within die delay variation is measured by changing the supply voltage and temperature, using two identical ring oscillator chains (OSC) each with 95 NAND gates. The results are shown in Fig. 9, where X-axis show supply voltage and Y-axis show normalized delay variation at four temperature points. Normalized delay variation is calculated by taking the relative mean difference of measured delay between the pair of OSC, at each temperature and supply voltage point, and it is normalized with that of 1.2-V supply voltage at 25°C temperature. It can be seen that normalized delay variation is smallest at nominal supply (1.2-V) and room temperature (25°C), it increases by up to 15X at 79°C when supply voltage is reduced from 1.2-V to 0.5-V. This shows that at lower voltage and higher temperature, the effect of V_{th} variation has greater impact at oscillator frequency variation [15]. This means that due to process variation, the state integrity of

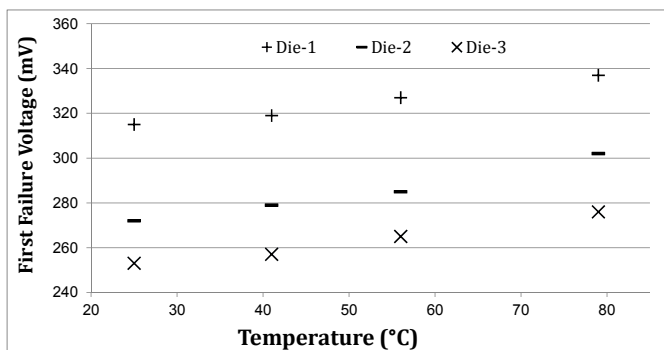


Fig. 8. Measured results showing the first failing voltage point of flip-flops due to within die temperature variation.

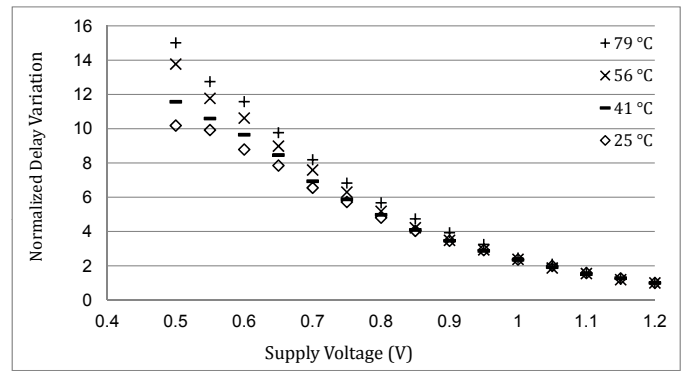


Fig. 9. Measured results of Intra-Die PVT Variation on delay of the test chip. These results demonstrate that due to change in temperature the effect of within die process variation gets worse as shown by within die higher normalized delay variation.

a flip-flop (Fig. 2) is more vulnerable at reduced supply voltage and higher temperature. Note that these results (Fig. 8 and Fig. 9) are specific to this technology library and are shown for illustration purposes only. For smaller geometry (below 65-nm) temperature spread and delay variation may be effected by for example additional mobility caused by increased mechanical stress and lower threshold voltage.

As shown in Fig. 8, First Failure Voltage (FFV) point of a flip-flop increases with increase in temperature. This means “Sleep State” voltage should take temperature variation into account to ensure state integrity. This has an effect on leakage power consumption of a design in “Sleep State”. To get an insight into voltage scaling and leakage power, Fig. 10 shows “Sleep State” leakage power by measuring I_{ds} , and varying the supply voltage after setting $V_{gs} = 0$. The measurements were carried out on a test chip (Die-2, Fig. 7) under 4 different temperature settings: 25°C, 41°C, 56°C, and 79°C. The x-axis shows the supply voltage ranging from 0.3-V to 1.2-V. The y-axis shows the normalized leakage power using log-scale. It can be observed that leakage power reduces exponentially with reduction in supply voltage, and 97.5% leakage power minimization is possible by reducing the supply voltage from 1.2-V to 0.3-V. The effect of temperature variation can also be observed, as can be seen, at a given voltage, leakage power increases with temperature. The leakage power at 79°C is an order of magnitude higher than at room temperature (25°C).

In this work, Minimum Retention Voltage (MRV) is defined as the scaled supply voltage value, at which all flip-flops in a given design can still preserve their state integrity. For a given technology and flip-flop design, minimum retention voltage of a design has to be characterized across all process and temperature corners to ensure state integrity. Through the trend shown in Fig. 7 (Sec. III-B) and Fig. 8 (Sec. III-C), we make an important observation. *Due to process variation, MRV varies from die-to-die and characterizing each die separately will not only ensure state integrity but can also minimize retention voltage per die, thus reducing leakage power.* In this case, MRV can be calculated by adding a voltage margin (referred as Retention Voltage Margin (RVM)) to First Failure Voltage (FFV) of the given die, that is

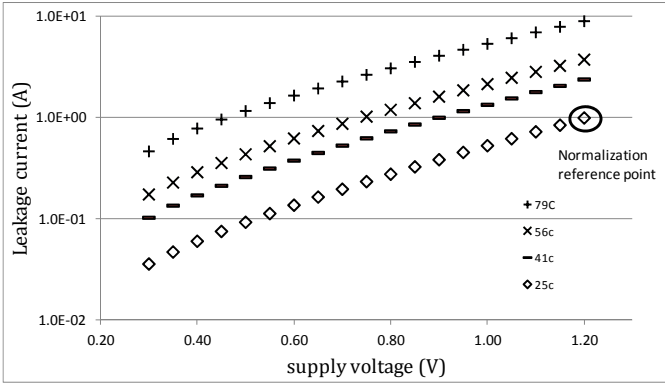


Fig. 10. Measured Test Chip Leakage Power Normalized to 1.2V nominal supply voltage at 25°C.

$MRV = FFV + RVM$. For example setting RVM to 54-mV for all dies (see Sec. IV-A for details of calculating RVM), the MRV of Die-1 is $315 + 54 = 369$ -mV, and that of Die-3 is $249 + 54 = 303$ -mV. Therefore setting MRV of individual die separately is beneficial to leakage power minimization, when compared to a technique that sets MRV of all dies using worst-case process and temperature corners. This observation is exploited in our proposed technique (Sec. IV), which employs a characterization algorithm to identify MRV of each die to minimize leakage power.

IV. PVT AWARE STATE PROTECTION TECHNIQUE

PVT variation analysis discussed in the previous section show two important observations. Firstly, 79% of all dies exhibit single bit failure at FFV, while the rest show multi-bit failure. Secondly, MRV characterization per die is beneficial to leakage power minimization. These two observations are used to develop a simple and effective technique to improve state-integrity of voltage scaled flip-flop under process, voltage and temperature variation. The proposed technique consists of the following two steps. Firstly, a characterization algorithm is used to determine MRV of a given die, this is because MRV of each die varies due to process variation as observed in Fig. 7. The characterization step of a die is an offline process and is performed only once per die. Secondly, a control flow for error detection and single-bit error correction is proposed, which relies on horizontal and vertical parity; whenever an error is detected, it raises the characterized minimum retention voltage to reduce subsequent error possibility. The prototype of the proposed control flow is implemented in the host computer using Python script, which provides voltage scaling by controlling an external power supply to the test chip (Fig. 1-(b)).

A. Characterization Algorithm

For each die, First Failure Voltage (FFV), is determined through voltage scaling, and then a Retention Voltage Margin (RVM) is added to FFV to get the minimum retention voltage (MRV) of each die that is $MRV = FFV + RVM$. The added Retention Voltage Margin (RVM), is the sum of Temperature Variation Margin (TVM) and Safety Margin (SM).

Input: Initial Retention Voltage (IRV), Voltage Scaling Resolution (VSR), Retention Voltage Margin (RVM)

Output: Minimum Retention Voltage (MRV)

- 1: $V_{correct} = IRV$; $V_{fail} = 0$
// $V_{correct}$ is the lower bound of supply voltage for correct state retention and V_{fail} is the upper bound of supply voltage for failed state retention
- 2: Current Supply Voltage = $\frac{V_{correct} + V_{fail}}{2}$
- 3: **while** $V_{correct} - V_{fail} > VSR$ **do**
- 4: **if** error detected **then**
- 5: $V_{fail} =$ Current Supply Voltage
- 6: **else**
- 7: $V_{correct} =$ Current Supply Voltage
- 8: **end if**
- 9: Current Supply Voltage = $\frac{V_{correct} + V_{fail}}{2}$
- 10: **end while**
- 11: $FFV = V_{fail}$
- 12: $MRV = FFV + RVM$
- 13: **return** MRV

Fig. 11. Process and Temperature Variation Aware Minimum Retention Voltage (MRV) Characterization Algorithm at 25°C.

Temperature Variation Margin is the worst case difference in FFV at the highest and the lowest operating temperatures, for a given technology and flip-flop design when considering process variation. In this work, TVM is set to 30-mV by using the maximum FFV difference of three corner case dies (Die-2; Fig. 7), as shown in Fig. 8. Safety Margin is set to 2% of nominal supply voltage. In this work nominal supply voltage is 1.2-V, and therefore safety margin is set to 24-mV. Therefore, retention voltage margin (RVM) is set to $30 + 24 = 54$ -mV.

For each one of the test chip, MRV is determined through a characterization algorithm (at room temperature 25°C) shown in Fig. 11. It requires three inputs: 1) Initial Retention Voltage (IRV), as a starting point to determine FFV, 2) Voltage Scaling Resolution (VSR), and 3) Retention Voltage Margin (RVM). IRV, VSR and RVM is determined through the following criteria: Fig. 7 shows measured results to determine the difference between first and second failure voltage points across all test chips. We used these measurements to set the value of IRV to 400-mV. This is because none of the dies fail at this voltage. The VSR is set to 1-mV, which is the smallest step size supported by the external power supply source (Agilent U3606A). Finally, RVM (Retention Voltage Margin) is set to 54-mV to accommodate safety margin and effect of temperature variations.

As can be seen in Fig. 11, the algorithm starts by setting $V_{correct}$ to IRV, and V_{fail} to 0-V. $V_{correct}$ is the lower bound of supply voltage for correct state retention, and V_{fail} is the upper bound of supply voltage for failed state retention. Next, to determine First Failure Voltage (FFV), the algorithm reduces the difference $V_{correct}$ (correct state retention) and V_{fail} (failed state retention) by iterating until the difference between these two variables is smaller than VSR, which is minimum resolution of the power supply. In line-2 of the algorithm, the current supply voltage is set to the mid-point of $V_{correct}$ and V_{fail} . In each iteration, the two variables are

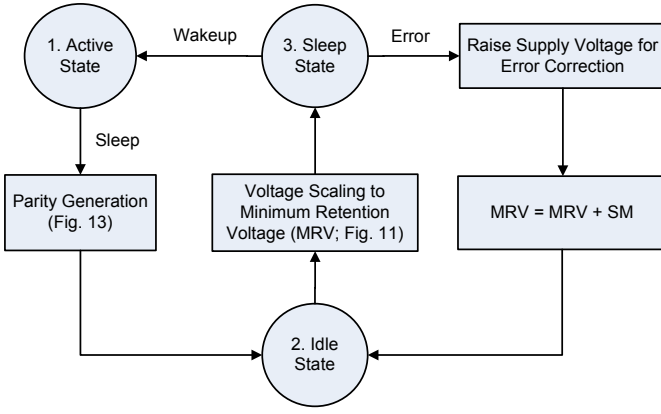


Fig. 12. Control Flow for State Monitoring and Protection of Flip-Flops for Voltage-Scaled State-Retention.

updated, if state corruption is detected, V_{fail} is raised to the current supply voltage, otherwise $V_{correct}$ is reduced to current supply voltage. This process is repeated by changing current supply voltage to the mid-point of updated $V_{correct}$ and V_{fail} . The loop exits with V_{fail} holding First Failure Voltage (FFV) value (line-11). Finally, the algorithm adds Retention Voltage Margin (RVM) to the observed FFV to calculate Minimum Retention Voltage (MRV) of the given test chip.

B. Control Flow

The control flow is implemented using a Python script running on host computer to communicate with the test chip through USB interface, and the test chip is powered by external power supply (Fig. 1-(b)). Fig. 12 shows the control flow of the proposed technique. It consists of three states: Active State, Idle State, and Sleep State. It can be seen that as soon as ‘sleep’ signal is received from the host computer during ‘Active State’, the parity is generated from the current flip-flop data and is stored in parity storage unit (Fig. 1-(a)). Parity generation and its storage is controlled by a micro-controller (ARM Cortex M0). This is why the micro-controller and parity storage unit is placed in always-on power domain (Fig. 13). Once parity is stored, the design goes to ‘Idle State’, after which the clock is stopped, the output of retention register block is isolated, and supply voltage is scaled down to pre-characterized Minimum Retention Voltage (MRV; Fig. 11). The design then goes to ‘Sleep State’. During ‘Sleep State’, the flip-flop states are continuously monitored and compared with the stored parity bits. In case of a mismatch, an ‘Error’ signal is generated in the form of hardware interrupt, which is received by the micro-controller. In response to that interrupt, the micro-controller raises the supply voltage to nominal supply voltage (1.2-V) and uses parity information (computed and saved) for single-bit error correction. In case error correction fails due to multi-bit errors, the control software is notified through USB interface. Software state recovery such as checkpointing can be used [16], however it is out of scope of this paper. In case of an error, the pre-computed Minimum Retention Voltage (MRV) is raised by Safety Margin (SM) which is set to 2% of nominal supply voltage (24-mV) to

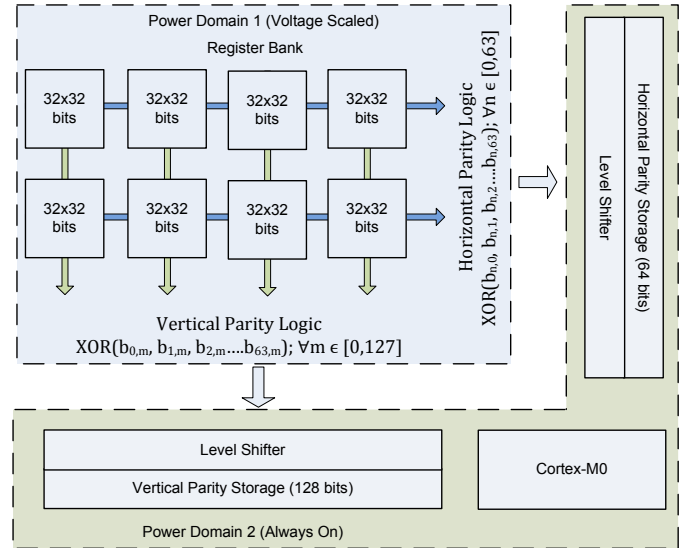


Fig. 13. Vertical and horizontal parity protected retention register block.

avoid subsequent errors. The updated value of MRV is stored in the host computer, which is used in subsequent ‘Sleep State’. After increasing the MRV, the control is transferred to ‘Idle State’, which in turn reduces the supply voltage to newly calculated MRV, and the design enters ‘Sleep State’. Finally, upon receiving a ‘wake-up’ request during ‘Sleep State’, the supply voltage is raised to nominal supply voltage, and the design enters ‘Active State’.

Fig. 13 shows the schematic of the retention register block that is protected using horizontal and vertical parity logic. The register block (Fig. 1-(a)) contains 8192 flip-flops, which are divided into 8 block, each with (32X32) 1024 flip-flops. The control of the parity logic is provided by ARM Cortex-M0 micro-controller, which is a 32-bit 3-stages pipeline RISC processor. There are two power domains (PD) in the design (Fig. 13). Power Domain 1 (PD-1) is used for register block, which can be scaled down during state retention mode through external Power Supply. Power Domain 2 (PD-2) is used for parity storage and micro-controller, which is kept in always-on power domain (always operating at nominal supply voltage of 1.2-V) for continuous state monitoring of register block.

C. Design Synthesis Flow

Fig. 14 shows the design synthesis flow which incorporates the horizontal and vertical parity insertion. In a conventional digital circuit design, firstly the RTL of a circuit is converted to gate level netlist through logic synthesis, which is followed by scan chain insertion for manufacturing test. The last stage is placement and routing. For the proposed design flow, an additional step is needed after scan chain insertion for horizontal and vertical parity insertion. This is because the scan chain converts distributed flip-flops to structured arrays. A tcl script is used to read the output of DFT tool, after scan chains have been inserted, and it connects all flip-flops along the scan chains for horizontal parity generation. Similarly for vertical parity generation, all flip-flops at the same depth of each scan chain are connected together. This concept is

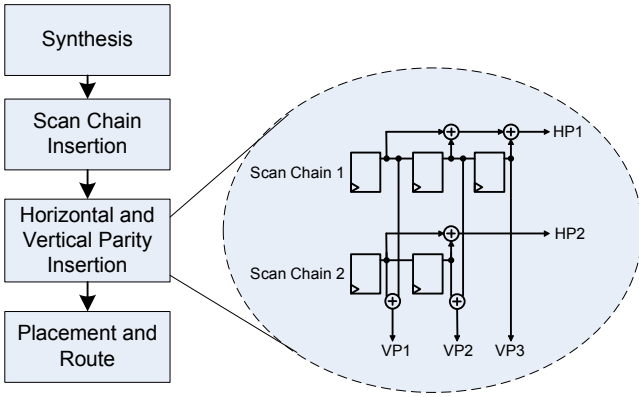


Fig. 14. Vertical and horizontal parity insertion design synthesis flow.

elaborated in Fig. 14, which shows how flip-flop are connected for horizontal and vertical parity generation. Scan chains may have different number of flip-flops, in which case the missing flip-flop (Horizontally or Vertically) is replaced by using a direct connection. An example with two scan chains is shown in Fig. 14, where the first scan chains has three flip-flops and the second scan chain has two flip-flops. It can be seen that the first horizontal parity is generated by using two XOR gates, while the second horizontal parity is generated by using only one XOR gate. Likewise, the last vertical parity is generated without using any XOR gate.

1) *Implementation Cost:* We analyzed the overhead of the proposed technique in terms of area, power and delay. The proposed technique requires two XOR gates per flip-flop. This implementation has 8192 flip-flops and the parity logic is about 51% of the flip-flop area. The parity logic used in this work incurs about 2.7 additional nets per flip-flop. From Fig. 13, it can be observed that the number of horizontal parity storage registers is equal to the number of scan chains and the number of vertical parity storage registers is equal to the depth of the longest scan chain in the design. One level-shifter is needed for each of the parity storage register. There is negligible increase in delay and dynamic power in normal mode of operation, this is because the parity logic is disabled during that mode. However, the leakage power increases, which is proportional to the area overhead.

D. Silicon Results

We conducted two experiments to demonstrate improved state integrity of flip-flops with aggressive supply voltage scaling in “Sleep State” that is possible through the proposed technique. First experiment demonstrates improved state integrity of flip-flops in “Sleep State”, and second experiment demonstrates the effect of aggressive supply voltage scaling on leakage power savings.

1) *Improved State Integrity:* We conducted an experiment using three dies (Fig. 7), the operating temperature was set to 79°C by using a temperature chamber. For this measurement, the test board was connected with a host computer through a USB interface, and Python script was used to communicate between the computer and the test board. For each die, we repeated the following five steps: 1) Voltage of the design was

TABLE I
MEASURED RESULTS FOR THREE SELECTED DIES SHOWN IN FIG. 7 AT 79°C IN “SLEEP STATE”.

Die #	FFV (mv)	MRV (mV)	# of Errors Detected	Proposed Technique Response
Die-1	315	369	0	No Error Detected
Die-2	285	339	0	No Error Detected
Die-3	250	304	0	No Error Detected

set to 1.2-V, 2) A single logic value (logic-1) was stored in all 8192 flip-flops, referred as *initial logic state*. This is because our experiments indicate that Logic-1 state retention is about 3-times more vulnerable to bit failure than logic-0. 3) Supply voltage was reduced to respective characterized Minimum Retention Voltage (MRV) of each die for 30-minutes, 4) Supply voltage was raised back to 1.2-V, 5) Flip-flop states were compared with the *initial logic state* to determine if bit failure has been observed. Results are shown in Table I, which shows First Failure Voltage (FFV) and Minimum Retention Voltage (MRV) of each die. Fourth column shows the number of errors observed in each die, and the last column shows the response of the proposed technique. For all three dies, no error was detected at MRV. It is important to note that when using conventional techniques (Canary [6] and open-loop [5]), flip-flop status is unknown. However, through this technique, it is possible to detect multi-bit errors and correct single bit error, thus it improves overall confidence on flip-flop state integrity at reduced supply voltage. When discussing measured results across 82-dies shown in Fig. 7, it was highlighted that 79.27% dies exhibited only single bit failure at FFV. This is why parity logic capable of single bit error correction is used in the proposed technique to improve state integrity at reduced supply voltage. At FFV, Multi-bit errors were observed in case of 20.73% dies, which can be detected through the proposed technique, and can be dealt with software check-point technique as explored in recent publications [16].

2) *Aggressive Voltage Scaling:* Table I shows that the difference in MRV of Die-1 and Die-3 is 65-mV, while still preserving state integrity. In comparison to using worst-case MRV across all dies, the improvement in MRV is different for each individual die. The proposed characterization algorithm (Fig. 11) achieves up to 17.64% improvement in MRV in the best case (Die-1 compared with Die-3). This improvement

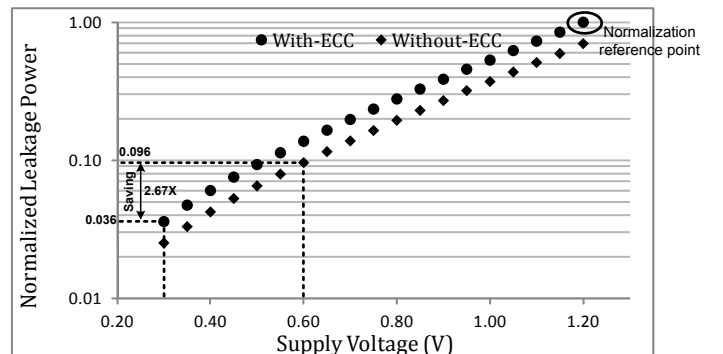


Fig. 15. Measured Leakage Power at room temperature normalized to 1.2 V supply voltage: With ECC vs. Without ECC.

is lower in the most common case (represented by Die-2), which is 8% lower (30-mV) than the worst-case MRV. To get an insight into potential leakage saving using this technique. We conducted an experiment with a design without ECC, which computes Minimum Retention Voltage (MRV) across all process and temperature corners and uses that single MRV across all dies. For example, in case of open-loop technique voltage is reduced to 0.6-V for all dies. On the other hand, the proposed technique employs a self-characterization algorithm (Fig. 11), which allows aggressive voltage scaling and each die has its own individual MRV. Fig. 15 shows normalized leakage power with and without using ECC. This measurement is taken at 25°C using Die-2 (Fig. 7). As can be seen, at a given voltage, the normalized leakage power of the proposed technique is higher than that of a design without ECC. This is because of 33% area overhead of parity logic. However, the proposed technique is capable of state retention at much lower voltage leading to overall lower leakage power in “Sleep State”, due to using characterization algorithm (Fig. 11). For example, in comparison to a design without ECC and state retention at 0.6-V, the proposed technique can retain states at 339-mV (for Die-2) leading to 2.67-times additional leakage power savings.

V. CONCLUSION

This paper presents measured results from silicon to show that the state integrity of flip-flops is affected by process, voltage and temperature (PVT) variation. Through measurements of 82 test chips, each with 8192 flip-flops, implemented using 65-nm design library, we have shown that at 25°C, state integrity of a flip-flop is affected by process variation leading to spread of First Failure Voltage (FFV), from 245-mV to 315-mV, with 79% of total dies exhibiting single bit failure at FFV, while the rest show multi-bit failure. Furthermore, due to temperature variation, it is found that FFV increases by up to 30-mV with increase in temperature from 25°C to 79°C. The effect of process variation is also studied using a 45-nm technology node through Monte-Carlo simulation, when compared with 65-nm technology, it is found that the overall distribution trend remains the same, however the mean FFV has shifted to a higher voltage. The effect of PVT variation on state integrity of flip-flops is addressed through development of PVT-aware state protection technique that ensures state integrity, while minimizing state retention voltage per die. The proposed technique consists of characterization algorithm to determine minimum retention voltage (MRV) of each die, and employs horizontal and vertical parity for error detection and single bit error correction. In case of error detection, it dynamically adjusts MRV per die to avoid subsequent errors. Silicon results show that at characterized MRV, flip-flop state integrity is preserved, while achieving up to 17.6% reduction in retention voltage across 82-dies.

ACKNOWLEDGEMENTS

The authors would like to thank EuroPractice Mini-ASIC program for silicon fabrication and packaging, and Dr. Harry Oldham for his insightful comments.

REFERENCES

- [1] M. Keating, D. Flynn, R. Aitken, A. Gibbons, and K. Shi, *Low Power Methodology Manual for System-on-Chip Design*. Springer, 2007.
- [2] R. Krishnamurthy, A. Alvandpour, V. De, and S. Borkar, “High-performance and low-power challenges for sub-70nm microprocessor circuits,” in *CICC 2002*, 2002, pp. 125–128.
- [3] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*. John Wiley & Sons, 2007.
- [4] K. Kumagai, H. Iwaki, H. Yoshida, H. Suzuki, T. Yamada, and S. Kurosawa, “A novel powering-down scheme for low vt cmos circuits,” in *VLSI Circuits, 1998. Digest of Technical Papers. 1998 Symposium on*, jun 1998, pp. 44–45.
- [5] T. Enomoto, T. Oka, and H. Shikano, “A self-controllable voltage level (svl) circuit and its low-power high-speed cmos circuit applications,” *JSSC*, vol. 38, no. 7, pp. 1220–1226, 2003.
- [6] B. Calhoun and A. Chandrakasan, “Standby power reduction using dynamic voltage scaling and canary flip-flop structures,” *JSSC*, vol. 39, no. 9, pp. 1504–1511, 2004.
- [7] J. Wang and B. Calhoun, “Canary replica feedback for near-drv standby vdd scaling in a 90nm sram,” in *CICC 2007*, 2007, pp. 29–32.
- [8] —, “Techniques to extend canary-based standby scaling for srams to 45 nm and beyond,” *Solid-State Circuits, IEEE Journal of*, vol. 43, no. 11, pp. 2514–2523, nov. 2008.
- [9] A. Sasan, K. Amiri, H. Homayoun, A. M. Eltawil, and F. J. Kurdahi, “Variation trained drowsy cache (vtd-cache): A history trained variation aware drowsy cache for fine grain voltage scaling,” *TVLSI*, no. 4, pp. 630–642, April 2012.
- [10] A. Nourivand, A. J. Al-Khalili, and Y. Savaria, “Postsilicon tuning of standby supply voltage in srams to reduce yield losses due to parametric data-retention failure,” *TVLSI*, no. 1, pp. 29–41, Jan 2012.
- [11] A. Kumar, H. Qin, P. Ishwar, J. Rabaey, and K. Ramchandran, “Fundamental data retention limits in sram standby experimental results,” in *Quality Electronic Design, 2008. ISQED 2008. 9th International Symposium on*, march 2008, pp. 92–97.
- [12] ARM, *Cortex-M0 Technical Reference Manual*, ARM, 2009. [Online]. Available: <http://www.arm.com/products/processors/cortex-m/cortex-m0.php>
- [13] W. Zhao and et al., “Rigorous extraction of process variations for 65-nm CMOS design,” *Semi. Manufac., IEEE Trans.*, vol. 22, no. 1, pp. 196–203, Feb. 2009.
- [14] O. Oklahoma State Univ., Stillwater, “Osu freepdk,” 2013, <http://vlsiarch.ecen.okstate.edu/flows/>.
- [15] S. W. Sun and P. Tsui, “Limitation of cmos supply-voltage scaling by mosfet threshold-voltage variation,” *JSSC*, vol. 30, no. 8, pp. 947–949, 1995.
- [16] I. Koren and M. Krishna, *Fault-Tolerant Systems*. Morgan-Kaufman, 2007.



Sheng Yang received the B.E degree in Electronic Engineering from University of Southampton, UK, in 2008. He is currently working towards his Ph.D. degree in University of Southampton, UK.

In 2007 he held an internship with NXP modelling a data hub using systemC. In 2011 he held an internship with ARM investigating data integrity of flip-flops. His research interests include: low power and fault tolerance techniques, computer architecture and embedded processor system design.



Saqib Khurshed received the B.E. degree in Computer Engineering from NED University, Pakistan, in 2001, the M.Sc. degree in Computer Engineering from King Fahd University (KFUPM), Saudi Arabia, in 2004, and the Ph.D. degree in Electronics and Electrical Engineering from University of Southampton, U.K., in 2010.

From 2005 to 2007, he served as a Lecturer with KFUPM. Currently he is working as a Senior Research Fellow in the School of Electronics and Computer Science, University of Southampton. He is

interested in all issues related to design, test, reliability and yield improvement of low-power, high-performance, multi-core designs and 3D ICs.



Bashir M. Al-Hashimi (M'99-SM'01-F'09) received the B.Sc. degree (with 1st-class classification) in Electrical and Electronics Engineering from the University of Bath, UK, in 1984 and the Ph.D. degree from York University, UK, in 1989. Following this he worked in the microelectronics design industry and in 1999, he joined the School of Electronics and Computer Science, Southampton University, UK, where he is currently a Professor of Computer Engineering and Director of the Pervasive Systems Center. He is ARM Professor of Computer

Engineering, and Co-Director of the ARM-ECS research centre. He is Associate Dean (Research) of the Faculty of Physical and Applied Sciences, University of Southampton, UK. His research interests include methods, algorithms and design automation tools for low-power design and test of systems-on-chip and embedded computing systems.



David Flynn, a Fellow in R&D at ARM Ltd, has been with the company since 1991, specializing in System-on-Chip IP deployment and methodology. He is the original architect behind ARM's synthesizable CPU family and the AMBA on-chip interconnect standard. His current research focus is low-power system-level design. He holds a number of patents in on-chip bus, low power and embedded processing sub-system design and holds a BSc in Computer Science from Hatfield Polytechnic, UK and a Doctorate in Electronic Engineering from

Loughborough University, UK. He has served as a Visiting Professor with the Electronics and Computer Science Faculty at University of Southampton, UK since 2008.



Geoff V. Merrett (GSM'06-M'09) received the B.Eng. degree (first class honors) in electronic engineering and the Ph.D. degree from the University of Southampton, UK, in 2004 and 2009 respectively. His thesis was titled "Energy- and Information-Managed Wireless Sensor Networks: Modelling and Simulation."

He is currently a lecturer in electronic and software systems at the University of Southampton, where he pursues research into low-power embedded electronics, energy-aware wireless sensing and sensor networks. He has published over 40 papers in journals and refereed conference proceedings, and has served on a number of technical programme committees for conferences and events. He has co-edited a book on "Wireless Sensor Networks: Application-Centric Design (InTech, 2011)", and co-authored a chapter on "Wireless Devices and Sensor Networks" in the book "Energy Harvesting for Autonomous Systems (Artech House, 2010)".