Aging-Resilient Design of Pipelined Architectures using Novel Detection and Correction Circuits

Hamed Dadgour
Department of Electrical and Computer Engineering, University of California, Santa Barbara
hamed@ece.ucsb.edu

Kaustav Banerjee
Department of Electrical and Computer Engineering, University of California, Santa Barbara
kaustav@ece.ucsb.edu

Abstract—Time-dependent performance degradation due to transistor aging caused by mechanisms such as Negative Bias Temperature Instability (NBTI) and Hot Carrier Injection (HCI) is one of the most important reliability concerns for deep nano-scale regime VLSI circuits. Hence, aging-resilient design methodologies are necessary to address this issue in order to improve reliability, preferably with minimal impact on the area, power and performance. This work offers two major contributions to the aging-resilient circuit design methodology literature. First, it introduces a novel sensor circuit that can detect the aging of pipeline architectures by monitoring the arrival time of data signals at flip-flops. The area overhead of the proposed circuit is estimated to be less than 45% compared to that of previous approaches, which are over 95%. To ensure the accuracy of its operation, a comprehensive timing analysis is performed on the proposed circuit including the influence of process variations. As a second contribution, this work presents an innovative correction technique to reduce the probability of timing failures caused by aging. This method employs novel reconfigurable flip-flops, which operate as normal flip-flops as long as the circuit is fresh, but function as time-borrowing flip-flops once the circuit ages. This unique flip-flop design allows optimization on the trade-offs of timing and area by employing such a technique. It is shown via simulations that by employing the proposed design methodology, the probability of timing failures in the aged circuits can be reduced by as much as 10X for various benchmark circuits.

Keywords- Reliability; Fault-Tolerance; Aging; Timing Analysis; Negative/Positive Bias Temperature Instability (NBTI/PBTI); Process Variation; Diagnostics and Built-in tests.

I. INTRODUCTION

The design of robust circuits and systems becomes an extremely challenging task as critical dimensions of transistors scale down. These escalating reliability concerns have generated great interest in designing fault-tolerant architectures in the literature [1]. "Transistor Aging" is one of these reliability issues that significantly threaten the reliability of high performance nanoscale circuits. Transistor aging is referred to a gradual reduction in the current drive capability of CMOS devices, which results in the performance degradation of nanoscale VLSI circuits. There are a number of mechanisms that are responsible for transistors aging; however, in recent years, Negative/Positive Bias Temperature Instability (NBTI/PBTI) and, to some extent, Hot Carrier Injection (HCI) have been identified as the most dominant aging mechanisms. Most importantly, it is experimentally shown that the impact of aging is more pronounced for high-k/metal gate transistors [2].

Transistor aging, if not accounted for properly, can lead to timing violations and circuit failures. Over the past few years, several works have attempted to address these concerns at the fabrication [3]-[4], circuit [4]-[7] and system-level [8]-[9]. Other research works have focused on statistical modeling and design optimization to reduce the impact of such time-dependent variations [11]-[16]. There have also been numerous proposals for aging-tolerant circuit design, which attempt to improve reliability by reducing the probability of timing failures caused by aging.

One simple circuit design approach to reduce the impact of aging is gate-sizing, which introduces additional guard-banding that can compensate for the effect of performance degradation [17]. However, such an over-design approach is extremely costly in terms of power and area. Body biasing of transistors has also been proposed as a potential remedy for the degradation of circuit performance [18]. Other aging-resilient design techniques have been introduced that employ "aging monitoring sensors" to monitor and detect the degradation of device characteristics (such as ION current) [19]-[20]. These techniques can be followed by diagnostic approaches such as body biasing or supply voltage tuning to recover the circuit performance [4]-[7]. The ION current monitoring technique is effective if one can assume that the performance degradation of circuits occur at the same pace as that of the drain current of transistors. However, this is not generally the case because the performance degradation is a function of many design parameters that differ for various transistors within a circuit. For instance, the ION reduction depends on “activity factor” (the percentage of time that transistors are under the negative bias stress) (Fig. 1 (a)). Since activity factors for different devices are not identical, the reliability degradation of an entire circuit can not be projected based on screening the ION reduction of a single transistor. This fact is demonstrated through circuit simulations for three selected ISCAS85 benchmark circuits [21]. The results indicate that the performance of various circuits degrade at different paces under an identical aging scenario (Fig. 1 (b)). Therefore, approaches that rely on monitoring the ION degradation of transistors are not suitable. Hence, a more effective technique for screening the aging process is to monitor the delay of digital circuits. For instance, a circuit failure prediction method is proposed in [22] that relies on measuring the delay of logic blocks (which are separated by flip-flop stages). Although this circuit can accurately measure the degradation of circuits, it needs a considerable amount of hardware overhead.

In this work, an effective and at the same time, low-overhead and low-power circuit is proposed for the early detection of aging in nanoscale circuits. These sensors can be employed in pipeline architectures (as shown in Fig. 2) to monitor the aging of combinational logic blocks (CLBs). Furthermore, a novel correction scheme employing a time-borrowing (TB) solution is proposed that utilizes reconfigurable flip-flops referred to as Reconfig TB FFs. This method reduces the probability of timing failures. In the proposed techniques (both for monitoring and mitigating aging), only minor modifications to the design of original flip-flops are required and the rest of the digital circuit remains intact as will be discussed later in detail.

In summary, this paper has two major contributions:
1) It proposes an area- and power-efficient technique to detect circuit aging in pipeline architectures. The proposed method achieves this goal by monitoring the arrival time of signals from CLBs to flip-flops; thus, it does not require any modification to the original logic blocks. A comprehensive timing analysis considering the impact of process variations is performed to ensure the accurate operation of these circuits.
2) The area overhead of the proposed circuit is estimated to be less than 45% compared to over 95% required by previous works.

Fig. 1. Device and circuit performance degradation due to aging: (a) the threshold voltage increase due to NBTI depends on the activity factor, which is defined as TL/TC. Here, TL is the period of time that the PMOS device is under the negative bias and TC denotes the clock period. (b) Performances of various circuits degrade at different pace under identical device aging scenario. This suggests that there is little correlation between the rates of device and circuit degradations.

Fig. 2. The proposed NBTI-resilient pipeline architecture employs novel aging monitoring circuits (denoted as “sensor”) and reconfigurable time-borrowing flip-flops (referred to as “Reconfig TB”). Note that unlike middle stages, the first and last stages do not use time-borrowing flip-flops.
(2) A novel time-borrowing-based solution is introduced to mitigate the impact of aging once it is detected. This method requires minor modifications to flip-flops; once more logic blocks remain untouched. A complete timing analysis is conducted to evaluate the accuracy and efficiency of the introduced time-borrowing approach. It is shown that the proposed circuit can significantly reduce the probability of timing failures.

The rest of this paper is organized as follows. Section II provides a brief introduction to the aging mechanisms in nanoscale CMOS technologies. Section 0 discusses the aging detection circuit proposed in this work. Section IV introduces the novel correction technique to alleviate the impact of aging; this section also includes the detailed timing analysis of the proposed method. The last section concludes this paper.

II. AGING MECHANISMS IN NANOSCALE CMOS

The aging mechanisms gradually reduce the \( I_{ON} \) current of transistor over time, which results in the increased propagation delay of signals in circuits. Over a long period of time, the delay increase can result in timing violations in critical paths where propagation delay can exceed the clock period thereby causing the malfunctions of digital circuits. It is widely known that Bias Temperature Instability (BTI) and Hot Carrier Injection (HCI) are the two major physical phenomena that lead to transistor aging [2]. BTI in PMOS devices is referred to as NBTI, since the gates of PMOS are “negatively” biased with respect to the source. Similarly, BTI for NMOS devices is referred to as PBTI.

A. Bias Temperature Instability (BTI)

BTI phenomenon increases the threshold voltage of both NMOS and PMOS devices and consequently lowers their \( I_{ON} \) current. Since \( I_{ON} \) current degradation is more severe in the case of NBTI [2], PMOS devices are selected here to illustrate the impact of BTI. NBTI phenomenon is caused by interface states, which exist at the border of the silicon channel and the high-k dielectric (Fig. 3 (a)). Dangling silicon bonds, which are present at the imperfect interface between silicon and high-k dielectric are pacified by hydrogen atoms (which are abundant in the IC fabrication process). Under negative gate bias conditions, holes induced in the channel can break the weak Si–H bonds according to a chemical reaction shown in (1). Then, Hydrogen atoms released by broken silicon-hydrogen bonds diffuse into the high-k dielectric (Fig. 3 (b)). A small number of these Hydrogen atoms can even diffuse further into the metal gate where they become trapped permanently (Fig. 3 (c)). It should be noted that the NBTI phenomena is predominantly a reversible process, which means that once the negative bias is removed, the hydrogen atoms inside the dielectric can return to their original locations and the re-form silicon-hydrogen bonds. However, those hydrogen atoms that have diffused into the metal are not able to diffuse back into silicon. As a result, a small positive charge appears at the interface of silicon and the dielectric whenever the transistor is stressed by a negative bias. Because of this positive charge, the threshold voltage of the PMOS device gradually increases over time.

\[
\text{SiH} + h^+ \rightarrow \text{Si} + H
\]  

(1)

B. Hot Carrier Injection (HCI)

Hot Carrier Injection (HCI) is another important mechanism that contributes to the aging of devices in nanoscale technologies [2]. The term ‘hot carrier’ refers to either holes or electrons that gain high kinetic energy after being accelerated by a strong lateral electric field. Due to their high energy, hot carriers can get injected and trapped at the silicon-oxide interface layer forming a small space charge, which expands gradually over time as more charges are trapped (Fig. 4). This space charge causes device degradation by increasing its threshold voltage.

Fig. 4. HCI mechanism in NMOS device: electrons gain high kinetic energy and can get injected into the oxide layer.

III. CIRCUIT-LEVEL AGING DETECTION

The early detection of delay degradation is vital for aging-resistant design techniques as it assists in taking prompt actions in order to prevent timing failures. One simple approach for aging detection is based on constant monitoring of the \( I_{ON} \) current of a few representative transistors and comparing them with a reference current source. However, it can be shown that the reliability degradation of an entire circuit can not be projected based on screening the \( I_{ON} \) reduction of a single transistor (Fig. 1 (b)). Thus, a more accurate indicator of the aging is the delay degradation of CLBs (Fig. 2) instead of probing the aging of individual transistors. In order to achieve this goal, the arrival times of different data signals (which are outputs of CLBs) are monitored at the input of flip-flop stages (Fig. 2). The latter technique is more precise because different CLBs do not necessarily age at an identical pace even if their transistors degrade at the same rate.

Fig. 5. Pre-sampling-based aging detection circuit: (a) circuit-level schematic, (b) signal waveforms when input D arrives well before the clock (the fresh CLB) and (c) signal waveforms when the CLB is aged.

Fig. 6. Schematic of the proposed aging detection circuit.

Fig. 7. Signal waveforms corresponding to different nodes of circuit shown in Fig. 6: (a) Fresh circuit, (b) Aged circuit and (c) different variables used in the timing analysis of the aging sensor.

A simple approach to monitor the performance degradation of circuits is shown in Fig. 5 (a) where the output of the combinational logic is monitored using a delay element and an extra “pre-sample flip-flop” component. The pre-sampled data is then compared to the actual value captured by the original flip-flop using an XOR gate. Assuming that \( t_{DCLK} \) refers to the time difference between the arrival of data signal (D) and the rise of clock (CLK), and \( t_{DCLK} > t_{DCLK} \) which allows enough time for it to propagate to X before the rise of CLK. Since, in this case, the pre-sample flip-flop captures the same value as the other flip-flop does and hence, the output of the XOR gate remains ‘0’. However, if \( t_{DCLK} < t_{DCLK} \) due to aging (Fig. 5 (c)), D is no longer able to promptly propagate to X and hence, the two flip-fllops sample different logic values which triggers ‘1’ at the output of XOR gate. In this case, one can conclude that the circuit is going to have the timing failures in near future. Although this circuit (Fig. 5 (a)) can detect the aging of circuits, it requires duplication of each flip-flop and hence, results in high circuit overhead. Another aging detection circuit is proposed in [22], which employs combinational circuits instead of duplicating all flip-flops as described in the above method. However, the circuit proposed in [22] also requires considerable hardware overhead.

A. A Low-Overhead Aging Detection Circuit

To overcome these shortcomings, a novel low power and area overhead circuit is introduced in this work as shown in Fig. 6. This circuit
achieves lower area and power overhead by replacing the additional flip-flop and the XOR gate with a simple combinational logic circuitry and a NAND gate, respectively.

The proposed circuit is composed of two stages. Stage #1 generates a “voltage glitch” on its output node, F, if the circuit is aged; otherwise, F remains low. Since a glitch is not a valid logic value, Stage #2 is designed to convert the glitches on F to stable logic levels on OUT. The detailed operation of these two stages will be discussed in the following subsections. Note that since logic symbols are used to sketch the flip-flop and XOR gate, the circuit shown in Fig. 5 appears to be more compact than the one proposed in this work. However, the area-efficiency of the proposed circuit becomes clear when one considers the actual transistor-level implementations of these two sensors.

B. Voltage Glitch Generator

A schematic diagram presenting signal waveforms at different nodes of the proposed circuit is shown in Fig. 6 to demonstrate its basic operation. In this figure, it is assumed that the inputs of flip-flops (D) are connected to the outputs of CLBs of the pipeline architecture (as shown in Fig. 2). The output of Glitch Generator (F) becomes high only if all three inputs of the NAND gate (X, Y and Q) are ‘1’. When the CLB is fresh, its output (D) arrives well before the clock edge and as a result, output F remains low since at any given instant, at least one of the inputs of the NAND gate is low (Fig. 7 (a)). On the other hand, if signal D arrives late enough due to the aging of the CLB, there will be a time interval where all three inputs of the NAND gate are ‘1’ which generates a “voltage glitch” at the output node F (Fig. 7 (b)). It should be noted that since Y becomes low with a delay after the clock edge, F eventually becomes low, as well. This will provide a “time window” during which, F can be high. Outside this time window, the fluctuations on D cannot generate any undesired glitch on F.

C. Voltage Glitch Detector

When a voltage glitch at the output node (F) can not be considered as a valid logic value, it must be detected and converted to an appropriate logic signal. This is accomplished with a dynamic inverter (Fig. 6) where F is connected to the input of the dynamic gate. The internal node of the dynamic inverter is charged to ‘1’ during the pre-charge phase (CLK=’0’). In the evaluation phase (CLK=’1’), depending on the logic values of inputs, it can be either discharged or allowed to retain its high logic level. If the input of the gate, the internal node of the inverter will be discharged to ground which forces OUT to become ‘1’. Otherwise, if there is no glitch or it is not strong enough, OUT will maintain its original ‘0’ logic value.

D. Timing Analysis of the Proposed Sensor

Since the proposed circuit does not employ any flip-flop and uses the delayed version of data and clock signals, one has to perform a complete timing analysis to ensure its accurate operation. In this analysis, for simplicity, it is assumed that the sensor circuit does not age and hence, all of its parameters are treated as constant values. However, in all circuit simulations, this assumption is dropped and the effect of aging on the sensor circuits is also included.

More detailed signal waveforms at the different nodes of the proposed circuit (Fig. 6) are shown in Fig. 7 (c) where it is assumed that the CLB is aged. In this figure, tDCLK and tCLKQ are the propagation delays between nodes D and CLK, and intermediate nodes X and Y, respectively; tCLKX is referred to as the delay between the edge of clock and the output of flip-flop (Q) and tCLKQ is defined as the difference between the arrival of D and the rise of CLK (TCLK). Additionally, TCLK and TQ denote the arrival time of CLK and Q signals, respectively where TQ represents the fall time of X.

To perform the timing analysis, one should note that in order for OUT to become high, the inputs of the NAND gate in Fig. 6 (X, Y and Q) must turn ‘1’ after the rising edge of the clock (TCLK). This circuit is designed so that Y becomes low at TCLK + tCLKY, thus, it provides a time window (TCLK < TQ < TCLK + tCLKY) during which values of X and Q determine the logic level of F. When the CLB is aged, for accurate detection, X must remain ‘1’ while Q switches to ‘1’ during TCLK < TQ < TCLK + tCLKY time period. Thus, high-to-low transition of X (TQ) and low-to-high transition of Q (TQ) must occur during this interval. Therefore, one can obtain the following inequalities:

\[ T_{CLK} < T_X < T_{CLK} + t_{CLKY} \]  
\[ T_{CLK} < T_Q < T_{CLK} + t_{CLKY} \]

Furthermore, the voltage glitch on F must be at least \( \Delta_{th} \) wide to ensure the proper switching of the dynamic inverter of stage #2. Note that \( \Delta_{th} \) depends on the design of the dynamic inverter and is therefore a known parameter in this analysis. Since the width of output glitch on F is equal to the time interval when D, X and Y are simultaneously ‘1’ (TQ - TQ), one can obtain another inequality as shown by (4):

\[ T_X - T_Q > \Delta_{th} \]

Calculating the arrival time of X and Q in terms of other variables shown in Fig. 7 (c), we get:

\[ T_X = T_{CLK} - t_{DCLK} + t_{IN} \]  
\[ T_Q = T_{CLK} + t_{CLKQ} \]

By substituting (5)-(6) in (2)-(4), one can obtain five constraints as shown in (7)-(11). Note that (7) always holds true and (8) is true when (11) is satisfied. Therefore, for the accurate operation of the proposed circuit, its components must be designed so that only the timing constraints shown in (9)-(11) are satisfied.

\[ t_{CLKQ} > 0 \]  
\[ t_{DCLK} > 0 \]  
\[ T_{CLKX} > t_{DCLK} \]  
\[ T_{CLKQ} > t_{CLKX} \]  
\[ t_{IN} > t_{DCLK} + t_{CLKQ} \]

It should be noted that in the presence of process variations above variables will no longer be deterministic values and hence, they must be treated as probabilistic distributions. Therefore, Monte Carlo based simulations are performed to ensure that the proposed circuit still operates accurately under the impact of process variations. The results of these simulations are shown in Fig. 8 (a) where X-axix is the difference between the arrival of data signal and the edge of clock (tDCLK) and Y-axis represents the probability that the arrival of data signal triggers the output of sensor circuit indicating the aging of the circuit. For low-levels of process variations (\( \sigma_{Vth}=10\% \mu Vth \)), the proposed sensor circuit can accurately detect the aging with 94% precision with a single sample. Under more severe process variation scenarios (\( \sigma_{Vth}=20\% \mu Vth \)), the precision of detection reduces to 80%.

Therefore, in order to improve the accuracy, the sensor circuit must capture more samples. Fig. 8 (b) plots the accuracy of detection for various numbers of samples considering three process variation levels. It can be observed that by capturing sufficient number of samples (>20), one can precisely predict the degradation of circuits. For instance, if twenty samples are acquired, the accuracy of prediction is \( \approx 99.74\% \) even under the severe process variations (\( \sigma_{Vth}=20\% \mu Vth \)). It should be noted that the other sensor circuits proposed in the literature experience the same difficulty if the impact of process variations are considered properly.
Note that the device aging is an extremely slow process and can take several months before any noticeable device degradation can be detected. Therefore, the speed of detection and the number of samples that must be obtained to determine the aging of the circuit are not critical issues. Hence, while the proposed circuit requires several samples, it is capable of accurately detecting circuit performance degradations.

E. Simulation Results

The proposed circuit is simulated using Predictive Technology Models (PTM) [24] models for 45nm technology node. The simulation results are summarized in Fig. 9 for three different scenarios. The waveforms of inputs, outputs and important intermediate nodes corresponding to Fig. 6 are included in this figure. In the first case (Fig. 9 (a)), the output of combinational logic (D) arrives at the input of flip-flop at \( \approx 470\)ps and the clock signal rises approximately 350ps after it. The voltage levels on nodes F and OUT remain almost zero indicating that circuit is not aged. In the second simulation, D arrives at \( \approx 480\)ps generating a small glitch on node F. However, the duration and the voltage level of this spike is not adequate to discharge the internal node of the dynamic inverter and hence, OUT once more remains low. Finally, in the third case, D arrives at \( \approx 490\)ps creating a wide glitch, which is able to discharge the internal node of the dynamic inverter and hence, OUT becomes high. These simulations indicate that if the combinational circuit ages sufficiently, OUT will be eventually triggered.

F. Comparison with the Existing Work

The proposed sensor is more area- and power-efficient compared to the one presented in [22] the pre-sampling-based circuit shown in Fig. 5. The area and power overhead associated with these sensors are compared and results are shown in Fig. 10 (a). In this figure, “overhead” refers to any circuit component other than the original flip-flop that is added to the circuit to measure the late arrival of signals at the input of that flip-flop. As it can be observed, the proposed circuit has clear advantages in terms of having minimal design overheads. Area and power savings offered by the proposed sensor become even more significant when a large number of sensors are deployed in VLSI designs. Furthermore, the impact of various levels of process variations on these sensors is investigated in Fig. 10 (b). Here, it is assumed that the parameter fluctuations affect all transistors in the designs. One can observe that the detection error rates (number of faulty detections to total detections) are comparable for all three sensors.

![Fig. 10. (a) Design overhead for the proposed sensor in this work, the one introduced in [22] and pre-sampling-based circuit shown in Fig. 5. (b) Comparison between the detection error rates of these designs for various degrees of process variations.](image)

### IV. A NOVEL CORRECTION TECHNIQUE

Once the performance degradation due to aging is detected, proper actions should be taken in order to prevent timing failures. In this work, a novel adaptive “time-borrowing” approach is proposed to reduce the probability of timing violations in aged circuits. This is achieved by replacing the original flip-flops in the pipeline architectures with reconfigurable flip-flops (to be discussed later). In order to understand its basic operation, in the following subsections, a brief introduction on time-borrowing techniques is provided and then, the details of the proposed design are presented.

A. Timing Constraints in Conventional Pipelines

The accurate operation of a pipeline circuit depends on satisfying the Setup and Hold time requirements of flip-flops. Setup time by definition is the minimum time that the input of a flip-flop must be valid before the edge of the clock and Hold time is defined as the minimum time that input must be maintained valid after the edge of the clock. In pipeline architectures, the Setup and Hold time requirements can be calculated in terms of other circuit parameters as shown in Fig. 11. In this figure, \( t_{\text{_snap}} \) and \( t_{\text{hold}} \) denote the Setup and Hold time of flip-flops, respectively; \( t_{\text{CLB}} \) represents the latency of CLBs and \( t_{\text{CLKQ}} \) refers to the delay from the edge of the clock to the output of the flip-flop. The Setup time constraint can be expressed as:

\[
t_{\text{CLB}} + t_{\text{CLKQ}} + t_{\text{snap}} < t_{\text{C}} \quad (12)
\]

Assuming a constant clock period \( t_{\text{C}} \), the Setup time constraint determines the maximum allowable \( t_{\text{CLB}} \), beyond which, flip-flops of the next stage captures wrong logic values. The upper bound on \( t_{\text{CLB}} \) is:

\[
\text{Max}(t_{\text{CLB}}) = t_{\text{C}} - t_{\text{CLKQ}} - t_{\text{snap}} \quad (13)
\]

On the other hand, the Hold time constraint can be written as:

\[
t_{\text{CLB}} + t_{\text{CLKQ}} + t_{\text{hold}} \quad (14)
\]

Hold time imposes a limit on the minimum latency of the CLBs, below which, data signals can pass through two pipeline stages in a single clock cycle. This lower bound can be expressed as following:

\[
\text{Min}(t_{\text{CLB}}) = t_{\text{hold}} - t_{\text{CLKQ}} \quad (15)
\]

As it will be discussed in the following subsection, the upper and lower bounds on the latency of CLBs have important implications for the design of different pipelining techniques.

#### B. Timing Constraints in Time-Borrowing Pipelines

Time-borrowing in a pipeline structure allows logic blocks to automatically use slack time from a previous cycle [23]. It should be noted that in this technique, the transfer of slack time from one cycle or to the next does not require any additional circuitry or clock adjustments. The key advantage of this approach is that it allows logic blocks to use more than one clock cycle. Therefore, time-borrowing increases the maximum allowable delay of CLBs. This property of time-borrowing architectures can be utilized to reduce the probability of timing failures when CLBs age.

Time-borrowing pipelines are implemented by replacing conventional non-time-borrowing (NTB) flip-flops with their time-borrowing (TB) counterparts. To understand the behavior of a time-borrowing flip-flop, first, one should consider the temporal characteristics of a conventional flip-flop as shown in Fig. 12. In this graph, the X-axis is the delay from the input data signal (D) and the clock signal (CLK) which is denoted as \( t_{\text{DCLK}} \). The value of \( t_{\text{DCLK}} \) is negative/positive when the data arrives before/after the edge of the clock. The Y-axis is the propagation delay from the input (D) to the output of the flip-flop (Q) which is referred to as \( t_{\text{Q}} \). It can be observed from this graph that if D arrives sufficiently long before CLK (\( t_{\text{DCLK}} < -10\)ps), Q will capture an accurate logic level after \( t_{\text{Q}} \approx 50\)ps. However, for \( t_{\text{DCLK}} > -10\)ps, the output of the flip-flop fails to follow its input. Therefore, one can conclude that the Setup time for this flip-flop is \( \approx 10\)ps.

**Fig. 11. Setup and Hold time constraints in pipeline architectures:** (a) the maximum delay of each CLB is dictated by the Setup time; (b) the minimum latency is restricted by the Hold time requirement.

**Fig. 12. Timing diagram of a conventional non-time-borrowing (NTB) and a time-borrowing (TB) D-flip flop.**

Using time-borrowing flip-flops [23], one can shift the NTB curve shown in Fig. 12 toward right to make the positive values of \( t_{\text{DCLK}} \) also feasible (TB curve). This means that Q will capture the accurate value of D even if it arrives after the edge of the clock. The time-borrowing flip-flop allows a longer time interval for those signals, which arrive late at the input of flip-flop and thus, preventing timing failures.

Using time-borrowing pipelines requires careful timing analysis considering the Setup and Hold time requirements of flip-flops. Such requirements are summarized in (16)-(20) for a two-stage pipeline architecture shown in Fig. 13.

\[
t_{\text{CLB}} + t_{\text{CLKQ}} + t_{\text{SNAP}} < t_{\text{C}} + t_{\text{B}} \quad (16)
\]

\[
t_{\text{CLB}} + 2t_{\text{CLKQ}} + t_{\text{SNAP}} < t_{\text{C}} + t_{\text{B}} \quad (17)
\]

\[
t_{\text{CLB}} + t_{\text{CLKQ}} + 2t_{\text{SNAP}} < 2t_{\text{C}} + t_{\text{B}} \quad (18)
\]

\[
t_{\text{CLB}} + t_{\text{CLKQ}} > t_{\text{hold}} + t_{\text{B}} \quad (19)
\]
In these equations, $t_{CLB1}$ and $t_{CLB2}$ denote the propagation delay of signals through CLB1 and CLB2 (Fig. 13), respectively; $t_c$ and $t_b$ denote the clock period and the amount of time-borrowing, respectively. Above equations are similar to (12)-(15) except for the fact that in these calculations the amount of time-borrowing ($t_b$) is also incorporated.

\[
t_{CLB1} + t_{CLK} > t_{Hold} + t_b
\]  
(20)

Equations (16) and (17) express the basic setup time requirements for flip-flops, which state that the total time required for a signal to propagate from one flip-flop stage to the next (left side of the equation) must be less than the clock period ($t_c$) plus the borrowed time ($t_b$). Moreover, (18) states that the amount of time-borrowing ($t_b$) is shared between two stages. Thus, it indicates that the time-borrowing does not increase the effective clock period; it only allows one clock cycle to borrow a limited amount of time from the other cycles. Hence, the total time span available for a N-stage time-borrowing pipeline is equal to that of its conventional counterpart, the only difference being that the boundaries between different cycles are not defined as rigidly (Fig. 13). Similarly, equations (19)-(20) show constraints imposed by the Hold time requirements of flip-flops.

\[
\text{Max}(T_{CLB1}) + T_{CLB2} = 2t_c + t_b - 2t_{CLKQ} + 2t_{Hold}
\]  
(18)

\[
\text{Min}(T_{CLB}) = t_{Hold} + t_b - t_{CLKQ}
\]  
(19)

\[
\text{Min}(T_{CLB}) = t_{Hold} + t_b - t_{CLKQ}
\]  
(20)

\[
\text{Max}(T_{CLB}) = t_c + t_b - t_{CLKQ} + t_{Setup}
\]  
(16)

In this work, an adaptive time-borrowing approach (Adaptive TB) is proposed that can combine the advantages of both the conventional and time-borrowing pipelining techniques (Table 1). In this scheme, the circuit employs a conventional pipelining method before the aging of CLBs, which results in low Setup and Hold failure probabilities. After CLBs age (which can be detected by the proposed sensor), the pipeline system switches to a time-borrowing architecture, thereby resulting in low Setup and Hold failure probabilities. The proposed adaptive time-borrowing approach requires two components: (1) sensors for detecting the aging of CLBs (as proposed in the first part of this paper) and (2) reconfigurable flip-flops that enable the pipeline circuit to switch from the conventional architecture to the time-borrowing design.

D. A Novel Correction Technique Using Reconfigurable Flip-Flops

The proposed adaptive time-borrowing approach employs “reconfigurable flip-flops”, which operate as conventional flip-flops when the CLBs are fresh and perform as time-borrowing flip-flops when the CLBs are aged. The reconfigurable flip-flops are designed by introducing minor modifications to the design of traditional time-borrowing flip-flops. In order to understand how the reconfigurable flip-flops function, one needs to understand the basic operation of time-borrowing flip-flops.

In this work, the gate-level representations of the conventional, time-borrowing and reconfigurable time-borrowing flip-flops are shown in Fig. 16. These flip-flops employ identical master/slave latch architectures (Fig. 16 (a)). However, they differ in the clocking schemes that have been used to drive the clock signals of their master and slave latches (Fig. 16 (b)-(d)). In the conventional flip-flop (Fig. 16 (b)), when CLK = ‘0’ (CLK = ‘0’), the master latch is transparent; hence, the logic value of D appears at the input node of the slave latch. When CLK goes high (CL becomes ‘1’, as well), the slave latch becomes transparent, thereby, Q captures the logic value of D.

![Fig. 13. Schematic diagram to illustrate the timing constraints for the proposed time-borrowing pipeline architecture.](image)

![Fig. 14. Different constraints imposed by the setup and hold time requirement of flip-flops for a simple two-stage pipeline architecture.](image)

![Fig. 15. Comparison between the design spaces available for a two-stage pipeline using the conventional and the time-borrowing techniques; when the combinational circuits (CLBs) are (a) fresh and (b) aged. The dark solid circles represent the delay of various critical paths across the pipeline.](image)

![Table 1. Setup and Hold time failure probabilities for aged and fresh circuits using the conventional (Conv.), time-borrowing (TB) and adaptive time-borrowing (Adaptive TB) pipelining approaches.](table)
A time-borrowing flip-flop can be implemented by slightly modifying the clocking circuit of a traditional master-slave flip-flop (Fig. 16 (c)). The idea is to introduce an additional delay to the clock signal of the slave latch (two inverters highlighted in Fig. 16 (c)). The additional delay allows the data signal to propagate through the flip-flop even if it arrives shortly after the edge of the clock. The amount of time, which can be borrowed, is equal to the latency of the delay element.

The design of a reconfigurable flip-flop (Fig. 16 (d)) is similar to the one shown in Fig. 16 (c) except for an additional multiplexer component which allows the selection of the clock signal for the slave latch (Fig. 16 (d)). This multiplexer has two inputs which can be selected according to a shift register signal (NTBI). When the NBTI='0', the clock signal for the slave latch is connected to the inverted version of the CLK; hence, the circuit will operate as a conventional flip-flop (corresponding to NTB curve in Fig. 12). When NBTI='1', the clock of the slave latch is connected to the delayed version of the clock of the master latch. The delay caused by these inverters is equal to the total time-borrowing of the flip-flop (corresponding to TB curve in Fig. 12). In other words, the additional delay is the amount that the TB plot presented in Fig. 12 will be shifted to the right.

E. Simulation Results

Using circuit simulations, the effectiveness of the three pipelining techniques (conventional, time-borrowing and adaptive time-borrowing) is compared. A suitable metric is to evaluate the timing failure probabilities by identifying the Setup and Hold time violations. For instance, using Monte Carlo simulations of 10,000,000 two-stage pipeline architectures and assuming a moderate level of the process variation (σ=10% μVth), the probability of the Setup (Fig. 17 (a)) and the Hold time failures (Fig. 17 (b)) are calculated. In these simulations, it is assumed that the I_on of devices decrease by 10% when CLBs age. CLBs are chosen from different ICSCAS85 benchmark circuits. It can be observed that the conventional techniques results in a large number of Setup time violations (Fig. 17 (a)) when the CLBs are aged while the time-borrowing approach suffers from high Hold time failures (Fig. 17 (b)) when the CLBs are fresh. The proposed adaptive time-borrowing technique, however, effectively eliminates such timing failures in all cases. It can be observed that by using the proposed method, the failure probability is reduced by approximately one order of magnitude. Note that the results obtained from Fig. 17 are in agreement with the conclusions Table 1.

In order to investigate the impact of process variations, the effectiveness of three pipelining approaches are evaluated in terms of their failure probabilities (which are defined as the highest failure probability considering both the Setup and Hold time failures over the life-time of the circuit). The failure probabilities are shown in Fig. 18 (a) for various process variation levels (σ=10%, 15% and 20% μVth). It can be observed that the likelihoods of timing failures for all pipelining techniques increase approximately at the same rate by escalating the parameter fluctuations. Furthermore, the timing failure probabilities are investigated for deeper pipeline architecture with various numbers of pipeline stages as shown in Fig. 18 (b). It is shown that although the timing violations increase dramatically for the conventional and the time-borrowing techniques, the failure probability for the proposed approach increases at a smaller rate. Hence, the proposed technique can be more effective for deeper pipelines.

V. CONCLUSIONS

A novel aging-resilient design scheme is introduced to tackle the reliability concerns of nanoscale circuits caused by time-dependent performance degradation of circuits. To achieve this goal, unique circuit approaches are proposed for early detection of aging and mitigating its influence. This paper has two major contributions. First, a novel area- and power-efficient aging detection circuit is proposed. A comprehensive timing analysis of the proposed circuit is presented and its operation under the impact of process variations is investigated. Second, a novel correction technique is introduced by which one can reduce the probability of timing failures caused by aging. A complete timing analysis is performed to investigate the potential advantages of the proposed method. Simulation results indicate that this approach can reduce the probably of circuit failures by as much as 10X for various benchmark circuits.

ACKNOWLEDGMENT

This work was supported by Intel Corporation and the UC Discovery Program (Grant No. COM09S-156665).

REFERENCES