Analysis of Functional Errors Produced by Long-Term Workload-Dependent BTI Degradation in Ultra-Low Power Processors

Loris Duch,[1] Miguel Peón-Quirós,[1] Pieter Weckx,[2]
Alexandre Levisse,[1] Rubén Braojos,[3] Francky Catthoor,[2,4] David Atienza[1]

Abstract—Aging effects in digital circuits change the switching characteristics of their transistors, resulting in timing violations that can lead to functional errors at the system level. In particular, bias temperature instability (BTI) is a degradation effect that changes the threshold voltage of transistors. Its effect is more prevalent as the scaling of transistor dimensions progresses. In this work, we present a method to enable defect-centric long-term modeling of BTI degradation that takes into account the effects of concrete workloads at the processor data path level. Based on this study, we propose a novel design flow to link the impact of BTI degradation at the transistor (ΔV_{th}), processor data path (e.g., maximum frequency) and application-functionality levels. This flow may be used to improve system correctness over the entire device lifetime, avoiding unsafe working points, or to achieve a graceful degradation of system characteristics.

Our design flow is applicable to all types of digital circuits, including high-performance processors. However, in this specific work we focus on the domain of biosignal processing applications for wireless body sensor networks (WBSNs), whose pseudo-periodic nature interacts with the partially recoverable nature of BTI. Our results in this domain show, for a 32 nm implementation, a variation of up to 54.6 mV in the threshold voltage of the circuit transistors after one year of continuous operation, with an impact of 8.4 % in the maximum safe operating frequency. Such effects are expected to strongly worsen for longer lifetimes and more scaled technology nodes.

1 Introduction

Increasing the transistor density per chip, thanks to technology node shrinking over the years, has been the preferred solution to improve the performance of digital systems while reducing their energy consumption and production costs. However, that shrinking has increased the significance of effects such as process variation and device degradation, which are leading to increased reliability issues. Among those effects, bias temperature instability (BTI) is of key interest because of its dominant impact on timing variability in scaled devices and the complexity of its physical mechanisms.

In particular, BTI is a degradation effect that changes the threshold voltage (V_{th}) of transistors as charges get trapped at the channel/gate-oxide interface [1]: If the gate oxide traps gain enough energy, they may capture charge carriers, leading to a reduced amount of carriers in the channel and impacting the transistor V_{th}. Increased transistor integrations and clock frequencies lead to higher operation temperatures inside the chip. Since BTI has a strong dependency on operating temperature [2], its effects will become more pronounced with future down-scalings of transistors [3]. These variations in V_{th} affect the switching characteristics of the transistor, resulting in timing violations at the circuit level that can ultimately lead to functional errors at the system level.

BTI has two components: A recoverable part that disappears when the transistor is switched off, and a semi-permanent one that increases the extent of the previous as the circuit ages. The partially recoverable nature of BTI makes it dependent on the concrete circuit workload because that changes the duty cycle (DC) of each transistor in the network, that is, the time that the transistor is in direct polarization. Additionally, sleeping periods in systems executing pseudo-periodic applications enable a partial recovery of the switching characteristics. Therefore, it is necessary to know the concrete behavior of the applications that will run on the system to accurately predict the extent of BTI degradations on the long term and their impact on system-level functionality.

Having complete combinational circuits is also important to analyze the effects of BTI along signal propagation paths because it affects in different ways n- and p-type transistors, with cumulative effects that depend on their exact interconnections. Therefore, we study the impact of BTI at the level of a full processor ALU.

The main contributions of this work are the following:

- We present a framework for long-term extrapolation of defect-centric BTI modeling that preserves the characteristic behavior of each transistor for the concrete workload generated by an application, including sleeping periods.
- We use our framework to conduct a long-term study of the impact of BTI degradation at the processor data path level for lifetimes of up to ten years.
- We analyze how the effects of BTI aging at the level of individual transistors translate into data path level failures (e.g., timing variations in predicted critical paths that make a register load a value before the corresponding ALU output stabilizes). The flow works at a granularity of

[1] *Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland. Email: {loris.duch, miguel.peon, alexandre.levisse, david.atienza}@epfl.ch*
[2] *IMEC, Belgium. Email: {pieter.weckx, catthoor}@imec.be*
[3] *SMARTCardia, Switzerland. Rubén Braojos was with EPFL, Switzerland, during this work. Email: ruben.braojos@smartcardia.com*
[4] *Francky Catthoor is also with KULeuven, Belgium.*

individual processor instructions and/or operands.

- We analyze how these data path level failures translate into application-level functional errors (e.g., corrupted signal samplings or incorrect identification of ECG characteristics). We focus on the domain of wireless body sensor network (WBSN) applications because their typical pseudo-periodic behavior with significant idle phases is well suited to the study of partial recovery effects in BTI degradation. Additionally, these applications have well defined outputs (e.g., ECG sampling and feature extraction).

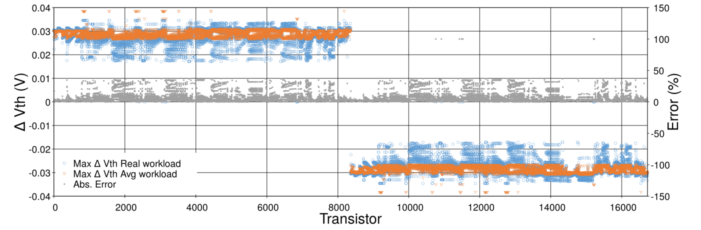
This work is organized in two parts. In the first one (Sections 3 and 4), we present an integrated analysis flow that combines BTI degradation modeling with workload-aware transistor-level dynamic timing analysis (DTA) to evaluate the extent of BTI degradation and its impact at the application functional level. This flow, based on work by Stamoulis *et al.* [4], uses a workload-aware defect-centric model [5] to study BTI-induced timing degradation. As the computational cost of running the model for complex circuits over long periods is prohibitive, we use an extrapolation technique to extend the results obtained for each transistor over a short period to arbitrarily-long periods—considering a real workload. We analyze the accuracy of our technique by comparing its predictions with the results obtained running the model for one year for a reduced number of transistors. The final outcome of the model is an evaluation of the impact of the BTI degradation on errors at the application functional level. In the second part (Section 5), we present the results obtained applying our flow to a biomedical application running on WBSNs. Finally, in Section 6, we draw the conclusions of this work.

2 Background and Related Work

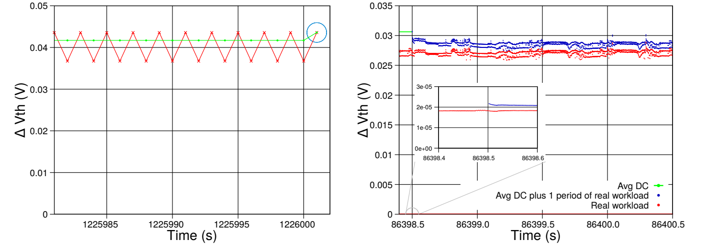
2.1 BTI degradation

BTI modeling. The exact physical causes of BTI have been the subject of long discussion. The reaction-diffusion (RD) model worked well for large transistors, because the random differences of individual defects average out, which enables the analytical models to represent accurately the effect of degradation along time and even at end-of-life (EOL) [6]. However, the down-scaling of transistors towards and beyond tens of nanometers reduces the number of defects per transistor responsible for time-dependent effects, making the stochastic nature of each defect and its impact on transistor characteristics more relevant [3]. This effect brought the validity of the RD model under question [7]. The defect-centric paradigm was introduced to study the contribution to transistor degradation of each individual defect through their individual carrier capture and emission time (CET), which can vary from µs to months [8]. The drawback of the defect-centric models is their high computational complexity, which limits their applicability to small sets of transistors or short-term characterizations. A good overview of both models is presented in [9]. The RD model has also been extended to cope with stochastic effects in reduced geometries, leading to the introduction of the double-interface RD model [10], [11]. However, contrary to the original RD model, it also suffers from high computational cost when calculating ΔV_{th} for AC stress, as it operates on a cycle-by-cycle basis [11].

In this work, we use the defect-centric model developed by Kaczer *et al.* [8] to obtain a characterization of all the transistors in the circuit during an initial period. We then use these results to extrapolate the impact of BTI degradation on ΔV_{th} for arbitrarily long periods with affordable computation times.



(a) Error introduced (grey) after 10 minutes of stress using averaged DC values (orange) instead of the real instantaneous DC (blue).



(b) BTI with average DC and one longest stress period at the end.

(c) V_{th} shift after average DC or real workload with long resting periods.

Fig. 1. Effect of DC, workload and resting periods on BTI degradation.

Full circuit coverage. BTI affects in different ways n- and p-type transistors. In particular, the effect on ΔV_{th}, although not symmetrical, is of opposite sign for each type. That means that the concrete interconnections of transistors in the circuit determine how n/p transistor degradations interact (compensating or accumulating) along the signal propagation paths. Hence, evaluating the degradations across the complete net instead of on individual transistors is essential to observe the real impact on the circuit critical paths. In contrast with previous works that study individual transistors or small benchmark circuits, we align ourselves with studies that tackle the complexity of large circuits representative of real systems [4], [12].

Workload-aware simulation framework. Along time, there has been some debate on whether BTI degradation depends on operating frequency [13] or not [6]. However, there is consensus in the link between the DC of each transistor, which is directly determined by the workload and the net structure, and the extent of BTI degradation. Kaczer et al. [2] propose that this link stems from the dual static/dynamic nature of the BTI effect: Under constant stress, BTI degradation worsens along time, but it sees an almost immediate recovery during relaxation periods. In that way, the dynamic switching of the transistor limits the extent of the degradation that would be seen under constant (static) stress. Furthermore, Santen et al. [14] show that instantaneous BTI has a strong local effect on transistor degradation, dominating the global trend at short scales. This combination of characteristics, which is specific to each particular transistor and workload, complicates modeling the behavior of the transistors using their average DC. In that line, our experiments confirm (Fig. 1a) that substituting the instantaneous DC of each transistor in our test circuit for its average DC introduces errors of up to 40 % after 10 min. This suggests that the concrete workload, which changes the length of individual stress and rest periods (not only their ratio), is relevant to accurately determine BTI-induced degradation along time.

To enable the processing of real workloads, Rodopoulos *et al.*

introduce a new signal representation, compact digital waveform (CDW), which aims to reduce the number of simulation steps required by atomistic models [15]. Their idea is to coalesce periods of input stimuli that have similar characteristics and run the model through them in a single step. However, their framework reuses identical stress patterns for every transistor, independently of their position within the netlist.

Following the previous observations, we propose a BTI analysis methodology that takes into account the characteristics of the real workload of the complete netlist to calculate the stress patterns of each transistor during its lifetime.

Long-term extrapolation. BTI degradation presents a typical logarithmic behavior along time with a ramp-up period during the first few seconds that corresponds to fast capture and release events [16], [17]. After longer periods (in the order of months to years), the degradation reaches a saturation point, which corresponds to the filling of the initial traps combined with a slower filling of the more energetically unfavorable (semi-permanent) traps. Due to these slower capture events, it becomes necessary to analyze the degradation over long periods to fully capture its impact on circuit characteristics, the relative order of critical paths and the possible appearance of functional errors during the system lifetime. However, defect-centric models struggle to reach long-term analysis because of their high computational complexity. This issue is exacerbated when full circuits are taken into account, as the switching activity of every node in the netlist has to be analyzed to detect possible changes in the critical paths along the complete period.

Stamoulis *et al.* introduced workload dependencies into CET-map based BTI modeling [4]. Their work can be potentially applied to any workload and circuit type. We build on their expertise by reusing the analysis flow for complete applications developed in their work. Among other major improvements to enable the discovery of functional errors, we introduce the possibility of accurately evaluating BTI-induced degradation over long periods of time. In comparison, to reach that goal they simply stretch the duration of each CDW point proportionally to cover the total desired period. This trivially enables the analysis over long periods without changing the computational complexity of the analysis. For example, with their approach, the processing of one data sample in periodic biomedical workloads can be stretched out along one year. Unfortunately, this is not the equivalent of simulating the system executing the application during that period, but of reducing the system working frequency by seven orders of magnitude. Indeed, for a periodic application that has 30 % of idle time at the end of each processing period, directly stretching the duration of the CDW points up to one year would produce an idle period of more than a hundred days at the end, which does clearly not represent its real working conditions.

A different approach is presented in [14], where the authors propose to calculate the average DC for each transistor and use it to extrapolate its aging until the desired period. Then, they find the longest stress period in the workload for that transistor and apply it after the desired aging, to calculate the highest combined long- and short-term degradation possible at that time. We compare their proposal with the effect of simulating a workload along all the analysis time in Fig. 1b. In the figure, we can see that their method (green line) matches almost perfectly the actual degradation for those conditions—a synthetic workload corresponding to the red line—as reported in their work. However, applying their technique to our real workloads produces much larger errors (e.g., larger than 10 %). The reason is that typical workloads in WBSN devices include long sleeping periods (e.g., while waiting for new data samples) that enable the release of most of the charges, so that the transistor starts *almost* from the initial condition in each working period. To verify our hypothesis, we aged a transistor during one day using its calculated average DC and afterwards we applied a complete workload period (2 s). Then, we calculated again the degradation with our BTI model using the real workload for the complete aging time. The results are shown in Fig. 1c: Simulating with the average DC predicts a larger long-term degradation, which leads to overestimation of the degradation during the final 2 s of interest. The root of this difference lies in the fact that the actual workload includes two stress periods (the two bands of each color in Fig. 1c) and a long relaxation period. This long recovery, which is absent in the simulation with average DC, enables almost complete relaxation (see inset in Fig. 1c), "pulling down" the degradation during the active periods. This experiment shows that, whereas BTI degradation can be approximated using the average DC for "classic" workloads that consist of continuous processing, new mechanisms are needed to accurately represent asymmetric workloads that include frequent long resting periods, as is typical in WBSNs running biomedical applications and in other types of event-driven Internet of things (IoT) devices.

In contrast with previous approaches, in Section 4 we propose a methodology that enables the use of defect-centric models over multi-year periods. Our method produces an efficient extrapolation of long-term BTI degradation that takes into account the concrete characteristics of the workload of each transistor in the circuit netlist at a dynamic granularity.

2.2 Evaluation of BTI-induced timing degradations

Stamoulis et al. designed a flow to analyze the impact of BTI degradation on the critical paths of a circuit [4]. In essence, their flow consists of the following steps: First, the workload at the circuit inputs is captured and propagated, through a flattened circuit structure, to all the transistors. To reduce the computational cost of computing with the BTI atomistic model over long signal traces, they identify regions of the workload with equivalent stress patterns that the model can process in a single iteration. For this, they use the decomposition of input stimuli into a CDW presented in [15]. Importantly, the length of the traces they analyze remains in the order of a few seconds. Then, they use a defect-centric BTI model [18] to determine the threshold voltage variation of each transistor; the obtained ΔV_{th} are then applied to each transistor in the netlist. Finally, they use Synopsys NanoTime [19] to perform static timing analysis (STA) on the circuit and derive the effect of BTI aging on the delay of each of the circuit combinational paths.

In this work, we extend their proposal as explained in Section 3 to introduce variable-length CDW decomposition, BTI-aging extension to arbitrary time lengths (even years for circuits with thousands of transistors), DTA via SPICE-like simulation of the original and aged circuits, behavioral simulation to obtain an error-free baseline output, and comparison of results to identify errors at the functional level.

2.3 Evaluation of functional errors

Beyond extending the framework for analysis of BTI-induced timing degradations at the circuit level, we also aim at evaluating their impact on the functionality of the complete system. In that

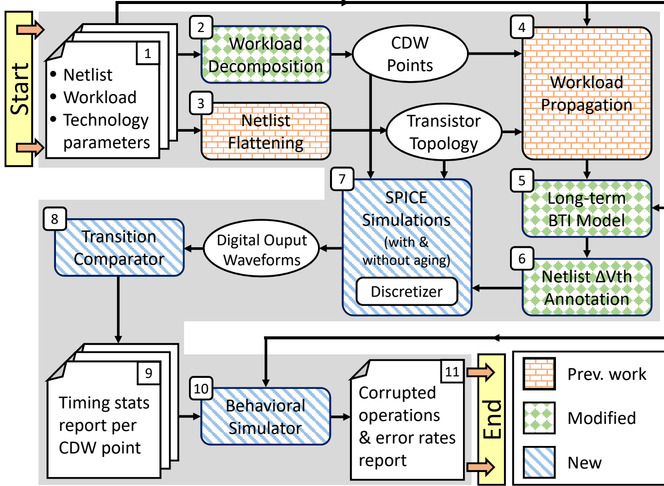


Fig. 2. Overview of our workload-aware flow for analysis of the impact of BTI-aging at the level of system functionality.

regard, Chen *et al.* proposed a methodology to model at the system level reliability degradations due to BTI effects in microprocessor architectures [20]. Their proposal is based on analyzing the delays on the most critical paths found with STA; hence, they give an estimation of the system lifetime, without providing information at the functional level, such as the type and amount of faulty operations generated by the processor pipeline. In contrast, our work uses DTA to observe and quantify the rate of BTI-induced functional errors on the circuit outputs over time. With the addition of a SystemC behavioral simulator, we can also identify which instructions will produce erroneous outputs. Furthermore, our framework enables the analysis of the impact of those errors on the quality of the results delivered by different biomedical applications running on the system.

3 Methodology for Evaluation of BTI Degradation Impact on Functional Errors

We extend the flow presented in [4] as detailed in Fig. 2. First, we introduce a variable-length decomposition of the workload into CDW points. Second, we modify the BTI evaluation step to characterize the effect of workload on each transistor, extrapolate it to the desired aging and then obtain the ΔV_{th} during one application period; this last period represents the working conditions of the application after the desired aging. Third, we perform DTA via SPICE-like simulation of the complete original and aged circuits. In comparison with previous works, our annotation provides individual aging conditions for each transistor in the netlist and propagates their concrete switching characteristics under workload. The use of DTA with BTI-aging annotation instead of STA produces an accurate characterization of the workload-dependent timing properties of each part of the circuit, hence identifying potential changes on the critical paths along the device lifetime. Finally, using a behavioral simulator to obtain an error-free baseline output, we can compare the results of each operation and of the complete application to quantify errors at the functional level and correlate them with circuit operations. The complete flow consists of eleven steps, as Fig. 2 shows:

1) Storage of the application workload in a value change dump (VCD) file as a succession of changes in the input signals of the circuit.
2) Workload decomposition into CDW points using the power gating signal of the processor as reference. Conversion into standard VEC file for use with SPICE during DTA.
3) Flattening of the circuit netlist for workload propagation and SPICE simulation.
4) Propagation of stress activity patterns across the netlist.
5) BTI modeling of defect activity based on CET maps using the real workload during the studied application period. We use the method explained in Section 4 to extend the BTI modeling time to arbitrary lengths while preserving the effects of workload and stochastic transistor variations. The result of this step is, for each CDW point in the workload, and for each transistor of the circuit, the ΔV_{th} produced by BTI after the desired aging period.
6) Update of the obtained workload-dependent ΔV_{th} shifts for each transistor of the flattened SPICE netlist and for each CDW point. This enables the execution of DTA to evaluate BTI degradation with workload dependency—in comparison with STA, which is workload independent.
7) SPICE simulations to evaluate the accumulated effects of BTI on the timing properties of the circuit (DTA).
8) Comparison of the binary values of the non-aged and the aged output signals obtained after the simulation for each CDW point using a transition comparator ("discretizer"). The discretizer takes into account the effects of data latching by the registers at the output of the considered circuits. This process establishes a link between BTI-induced timing degradations and the occurrence of functional errors when the propagated signal does not arrive before the next rising edge of the clock.
9) Elaboration of statistics for each CDW point and signal transition on the rising edge of the clock: slack time, propagation delay from input to output, time difference between the rising edge of the clock and the correct value of the signal when a timing violation occurs, etc. Every mismatch between the non-aged and the aged output signals on the rising edge of the clock is flagged to highlight possible timing violations.
10) Behavioral simulator to compare the outputs of the aged circuit with the correct ones and evaluate the impact of timing violations on its functionality. This simulator provides the "ground-truth" for the outputs that correspond to the operations in the application workload.
11) The final outcome is a report with the list of corrupted operations, their timestamps and their input operands.

The following paragraphs detail the most relevant steps of the flow, whereas Section 4 is entirely devoted to explain the modifications introduced to the BTI model to achieve long-term (up to ten years) analysis.

3.1 Variable size CDW decomposition

Typical biosignal processing applications have a pseudo-periodic workload composed of alternating active and idle periods to process input samples as they are acquired. Within one active period, the control and data signals of the processor (e.g., in the ALU) toggle according to the set of instructions executed and the values of the data variables; hence, the DC of each transistor is different for each active period. At a higher level, the DC of the application, characterized by the ratio between the active and idle (power-gated) periods, depends also on the sampling and

system operating frequencies. As a consequence, each transistor undergoes a particular combination of stress/relaxation during an active period, becoming fully relaxed during the application's idle times. Since the BTI effect is partially recoverable, the impact of these patterns is non-negligible; therefore, considering active and idle periods instead than just averaging the DC over the complete execution period is important to model accurately the recovery of the individual transistors [2].

The decomposition of the workload into a set of CDW points [4], [15] allows us to trade off between the computational effort and the accuracy of the BTI model: A higher number of CDW points increases the accuracy of the BTI-aware analysis at the cost of longer simulation time. The underlying idea in the CDW decomposition is that the model can calculate in one step the effect of BTI over a period of arbitrary length given a constant DC (and frequency). However, periods with different DCs generate different degradation (or recovery); hence, the BTI model should be (consecutively) run once for each one. Compared to previous works, we employ the power gating signal of the processor to identify the system idle periods and accurately create the variable-length CDW points. Each active period (i.e., computation burst) is uniformly subdivided into several CDW points, whereas a single point is used to characterize idle periods. In this way, we concentrate the effort of the BTI model on the most straining parts of the workload. As an added benefit, the CDW point decomposition enables the parallelization of the SPICE simulations during the evaluation of functional errors.

3.2 Transistor annotation throughout the circuit

In order to accurately analyze the degradation of each transistor, we propagate the stress activity patterns of the circuit—which can be characterized in terms of frequency f, DC α and duration Δt—across all the transistors of the netlist. To that end, we perform switch-level simulations of the circuit to compute the signal activities at each transistor, from which the f and α values are determined [4]. These parameters, which capture the V_{gs} stress voltage of each transistor, are subsequently provided as input for the BTI modeling stage. The simulations are performed based on the prior decomposition of the workload into CDW points.

3.3 Dynamic timing analysis

STA provides an approximation of the maximum operating frequency of a circuit. However, its limited accuracy under some scenarios leads in general to overly conservative results [21]. In contrast, DTA uses stimuli vectors to drive the activity of the circuit transistors in analog SPICE simulations, but DTA does not take into account circuit degradation due to factors such as BTI-induced effects. Therefore, significant guard bands are introduced in the maximum frequency to guarantee correct operation of the circuits [22], [23]. In our flow, we introduce aging-aware SPICE-based DTA analysis to evaluate the effects of BTI on the timing properties of the circuit without the need for worst-case guard bands. The division of the workload into CDW points enables easy parallelization of the SPICE simulations in multicore servers, running one CDW point simulation per core.

3.4 Analog signal discretization

To enable the comparison between the outputs of the non-aged and the aged circuits, we introduce a discretizer module that plays the role of a register interfaced between two consecutive processor pipeline stages. The discretizer makes the translation from the analog to the digital domains during every SPICE simulation, thus enabling the detection of timing violations that manifest as bit flips in register values. The module is implemented in Verilog-A and has no physical effect (e.g., parasitic load). To accurately determine the transition threshold values for our technology node, we analyzed a circuit composed of several flip-flops in series, toggling at a frequency of several GHz. At the input of the first flip-flop, we applied a slow trapezoidal signal with a very small slew rate. With this setup, we determined the transition threshold voltages *A* (0.399 V) and *B* (0.455 V) with an accuracy of $\pm 1\,\mathrm{mV}$ (Fig. 3), which is thus adequate in comparison with the normal variability margins considered in circuit design.

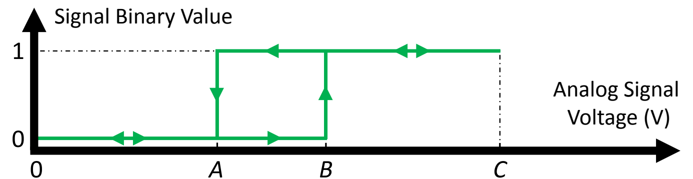


Fig. 3. Signal conversion from the analog to the digital domain with hysteresis. *A)* high-to-low ('1' to '0') transition threshold; *B)* low-to-high ('0' to '1') transition threshold and *C)*, nominal supply voltage.

3.5 Behavioral simulator

Timing violations caused by BTI aging at the circuit level do not necessarily translate into functional errors at the application level of the system. For example, a timing violation may occur on a combinational signal that is not part of the result of the current operation. With these considerations in mind, we have integrated a behavioral simulator (written in C++) in the flow to evaluate the impact of timing errors on the quality of the results produced by the application using high-level metrics such as the signal-to-noise ratio (SNR). The simulator helps also to uncover which operations on which exact functional units over concrete bits are the most likely culprits for the errors. That knowledge enables the introduction of high-level measures to control the quality of the delivered results (such as lowering the operating frequency, increasing temporarily the supply voltage or alternating between spare circuits) or the reinforcement of the affected circuits at the technology or architecture levels. In the case that a high-level (e.g., C++) behavioral simulator of the system under study is not available, an RTL model can be employed instead, albeit at a possible longer execution time.

4 Methodology extension for long-term BTI evaluation

The defect-centric model takes into account both the contribution of each individual trap event through CET maps, and the real workload of each transistor. Thus, it enables very precise simulations of the impact of BTI degradation on the switching characteristics of each individual transistor. However, its computational complexity is prohibitive to analyze large circuits over long time scales: Running the BTI model for a period of one year on just a single transistor requires around 5.3 days in an Intel Xeon Gold 6154 (3.0 GHz) server. Similar considerations apply as well to the double-interface RD model with AC stress over long periods.

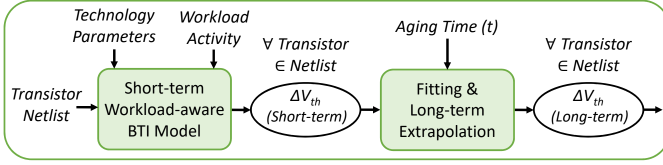


Fig. 4. Proposed long-term analysis methodology.

In this section, we propose a method to overcome the previous limitation and we assess the accuracy of its predictions by actually running the model to simulate one year of aging for a selected set of transistors. In essence, we create a curve that captures the overall logarithmic trend of the ΔV_{th} for each specific transistor ("DC curve"), and then we apply on it an "AC offset" that represents the exact conditions for one application period, at the desired point in time. Similar approaches have been employed before successfully [11], [14]. With this method, we can extend the BTI simulation to periods of months or even years with a small loss of accuracy.

4.1 Long-term extrapolation

Our long-term BTI modeling methodology to evaluate the impact of real (non-averaged) workloads for periods lasting up to several years within reasonable computing times consists of the following steps:

1) A workload-dependent BTI simulation is performed with the defect-centric model on all the transistors for a short aging period (e.g., 10 minutes). The simulation is executed for each transistor even if they have the same DC to capture the stochastic nature of individual trap and release events.
2) The time-dependent evolution of the ΔV_{th} curve obtained for each transistor is fitted to (1), thus capturing its logarithmic evolution along time (t):

$$\Delta V_{th}^{fitting}(t) = a \times (\log_{10}(t))^{b} + c \tag{1}$$

3) The obtained coefficients a, b and c are used in (2) to extend the ΔV_{th} curve of the transistor until the desired aging period (e.g., 1 year):

$$\begin{aligned}\Delta V_{th}^{extrapol}(t) = \left(a + \frac{a \times t}{\Phi}\right) \times (log_{10}(t))^{(b+(b\times t)/\Psi)} \\ + \left(c + \frac{c \times t}{\Upsilon}\right)\end{aligned} \tag{2}$$

4) Using the calculated ΔV_{th} for each transistor, a final BTI pass is performed during a last application period (e.g., 2 s), producing the concrete V_{th} shifts for each CDW point in that period.
5) The updated V_{th} values can be used in a subsequent DTA phase to evaluate the timing and functionality of the circuit after the desired aging period.

The fitting coefficients a, b and c of (1) are determined for each transistor independently based on the initial BTI simulation. The duration of the initial simulation (e.g., 10 minutes of aging in our case) must be long enough to capture the overall trend of the ΔV_{th} curve—well past the initial "ramp-up" time of the BTI effect. In our experiments, we produce independent fitting curves for CDW points with large behavior differences, separating those following a resting period from those in the middle of an active period. We introduce three calibration factors, Φ, Ψ and Υ to improve the quality of the extrapolation. Their values, which have been determined empirically, are identical for all the transistors of the circuit. Our methodology enables workload-aware analysis of BTI aging with a defect-centric model over arbitrarily long periods while limiting the modeling effort to a constant value—in our case, 10 min and 2 s of application time.

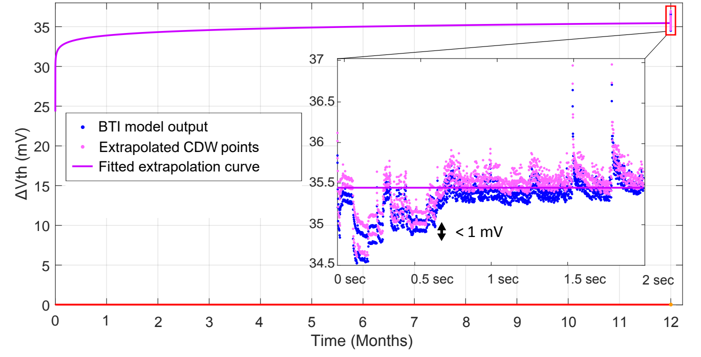


Fig. 5. Fitting extended to 1 year, with final superimposed offset corresponding to one application period of 2 s.

4.2 Validation

We carried out a study to validate the proposed methodology for long-term transistor-level BTI modeling using the 8 most active and 8 least active transistors of the ALU16 circuit, plus 8 more randomly selected. As a workload, we considered a set of 3000 CDW points covering 2 s of execution time, which corresponds to one ECG window (1000 samples captured at 500 Hz). This set of CDW points was repeated up to the desired time (i.e., to simulate an application repeating the same cycle continuously). This produces a more representative workload than simply stretching the duration of each CDW point while reducing the length of the stimuli traces.

First, we applied our extrapolation technique to extend a 10 min simulation up to one year. Then, we applied the full defect-centric BTI model during the same complete period to obtain an accurate reference of the BTI-induced aging (this step required 5.3 days per transistor, but it is not necessary for normal use of our methodology). Figure 5 shows the quality of the fitting and extrapolation for one of the selected NMOS transistors. The zoom over the last 2 seconds following 1 year of aging shows the instantaneous V_{th} values for the transistor after each CDW point calculated with our technique (pink points) and with the full model (blue points). By superimposing the series of points obtained with each method, we can observe that our technique introduces a deviation of up to 0.15 mV, which represents 0.41 % of the maximum ΔV_{th}. Similar results are obtained for the other transistors in this validation study, which present a maximum absolute deviation < 1 mV. These deviations are in practice negligible—especially taking into account the stochastic nature of defect-centric BTI modeling. Given the large gain achieved in execution time in comparison to applying the full BTI modeling to the entire multi-year period, while still keeping an accurate workload representation that enables analysis of instantaneous BTI effects, our approach fully solves the accuracy-execution time trade-off problem. Therefore, we conclude that our technique is suitable to conduct the needed long-term experiments for BTI-induced degradation exploration in the following section.

TABLE 1
Benchmark circuits from a 16-bit pipeline execution stage. Maximum operating frequency as determined with STA (at slack time = 0 ps).

Circuit	Description	Gates	Transistors	Freq. (MHz)	Period (ps)
Adder16	16-bit adder w/ carry	65	724	1645	608
Mult16	16-bit comb. mult.	1276	13 270	562	1778
MAC8	8-bit comb. mult. + 16-bit adder	1368	14 046	555	1803
ALU16	16-bit Arithmetic and Logic Unit	1672	16 704	552	1811

TABLE 2
Comparison of the analysis methods used in our experiments.

Technique	Workload	Accumulated aging	Ongoing aging
STA	No	No	No
DTA	Yes (SPICE)	No	No
DTA+BTI	Yes (SPICE)	Yes (BTI + (1), (2))	Yes (e.g., 2 s)

5 Experimental setup and results

5.1 Experimental setup

In this section, we explain the details of the hardware platform and the software application used in our experiments.

5.1.1 Biomedical application

To analyze the effects of BTI-induced degradation at different levels in a WBSN platform, we selected 3L-MMD [24], a complex single-threaded real-time biosignal processing application that executes pseudo-periodic tasks within fixed time boundaries under a tight energy budget. The pseudo-periodic nature of this workload is an interesting characteristic for our study: In contrast to standard test benches, it fosters the partial recovery of BTI effects after each active period, which may lead to reduced V_{th} degradations.

3L-MMD is a cardiac monitoring application that performs three-lead ECG delineation using multi-scale morphological derivatives (MMD) [25]. The first step of the application is a three-lead morphological filtering (MF) algorithm that removes artifacts (created by muscle activity, AC supply interferences and breathing-induced base drift) from an ECG acquisition [26]. The second step merges the filtered streams through a root-mean-square (RMS) combination. Finally, the third step performs the delineation of the ECG fiducial points. We considered an execution window of 1000 samples acquired at 500 Hz over 2 s. The activity traces were generated with a cycle-accurate SystemC simulator that implements the TamaRISC processor [27] and divided into 3000 CDW points.

To evaluate BTI-induced degradation on the application output, the ECG fiducial points are classified as correct, misplaced or missing. An ECG fiducial point is considered as correct when it is present in the heartbeat in the right sequence and within the correct time ranges. Present points with correct timing, but out of sequence, are considered as misplaced. Finally, points that are absent or out of the allowed time range are classified as missing.

In addition, we introduce two more metrics to evaluate partial degradations of the application output. First, the average time deviation (with respect to the error-free execution) of the ECG fiducial points that are classified as correct. Second, the percentage root-mean-square difference (PRD) is used to assess the diagnostic quality of compressed ECG records. We use the classification presented in [28] (very good: $0\,\%$ to $2\,\%$; good: $2\,\%$ to $9\,\%$; poor: $> 9\,\%$) to evaluate the maximum frequency that allows the system to produce biomedical results with medical significance.

5.1.2 WBSN hardware platform

Based on the chosen application, we extracted the ALU component from the execution stage of the pipeline of an ultra-low power WBSN [29]. The selected complete circuit, ALU16, performs signed and unsigned integer arithmetic and logic operations widely used in digital processing of biomedical signals. Therefore, it combines circuits for addition/subtraction, $16 \times 16 \rightarrow 32$ bit multiplication, multiply-accumulate, and additional logic operations in a 16-bit data path (biosignal samples are usually encoded on 16 bits [25], [30]). To enable more detailed analyses, we also study individually some of the most relevant components of the ALU: Adder16, Mult16 and MAC8, as listed in Table 1.[1]

Each circuit was synthesized with Synopsys Design Compiler with a constraint for minimum area, and using a 32 nm high-V_{th} and slow-slow (SS) process corner technology library. The parameters of the transistors were defined by the 32 nm predictive technology model (PTM), including a default V_{th} of –450 mV for PMOS transistors and 508.8 mV for NMOS transistors. As initial reference, Table 1 shows the maximum frequency determined through STA with Synopsys NanoTime for each benchmark circuit. For the BTI aging studies, we assume an internal temperature of 353 K and nominal supply voltage (1.05 V).

5.1.3 Experimental methodology

During our experiments, we use three different analysis methods. First, STA, which is aging and workload-agnostic, using Synopsys NanoTime. Then, SPICE-based (Synopsys CustomSim) DTA, which is workload-aware, but without considering any aging. Finally, DTA with accumulated and ongoing BTI-induced aging. Table 2 summarizes the characteristics of each analysis method.

For the aging-aware DTA analysis, we first perform our long-term extrapolation to calculate the ΔV_{th} at the start of the period of interest: We execute the BTI model for 10 min, fit the obtained curve using (1), and then calculate the long-term extrapolation using (2). This process is repeated for every transistor to capture the stochastic nature of individual trap and release events. Then, we run the BTI model for 2 additional seconds, recording the ΔV_{th} at the end of each CDW point. Finally, we use these ΔV_{th} values to conduct the SPICE simulations over the complete circuit.

Our method exploits long-term extrapolation to account for the previously accumulated aging while considering the characteristics of each transistor but, in contrast with previous long-term analysis methods, resumes a fully detailed BTI modeling to calculate the effect of aging under a real workload and circuit structure at the desired time. Therefore, it enables the detection of timing failures and, with the help of the simulator, pointing exactly to the processor operation that will likely produce that failure.

5.2 Results

In this section, we evaluate the effects of BTI-induced degradation at the circuit and application functionality levels.

1. In order to optimize the design area and to obtain a consistent comparison, the MAC unit in the full ALU16 unit reuses the 16-bit multiplier with an input operand size of 8 bits and considering only 16 bits from its output.

TABLE 3
Maximum transistor ΔV_{th} measured for the ALU16 after 2 s of circuit operation with varying accumulated BTI aging.

Aging	PMOS max $\lvert\Delta V_{th}\rvert$ (mV)	PMOS max $\lvert\Delta V_{th}\rvert$ (%)	NMOS max $\lvert\Delta V_{th}\rvert$ (mV)	NMOS max $\lvert\Delta V_{th}\rvert$ (%)
t = 0 s	37.1	8.1	36.3	7.3
t = 1 year	54.6	12.1	48.1	9.5
t = 10 years	**58.3**	13.0	**51.4**	10.1

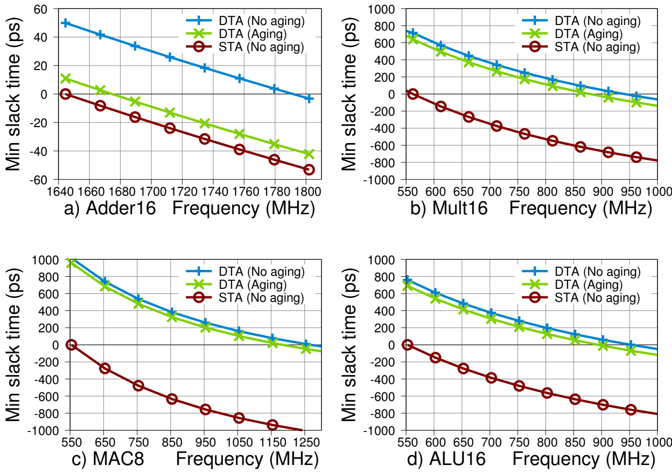


Fig. 6. Slack time at different working frequencies for each benchmark circuit, measured after 2 s of operation. The X axis (slack time = 0 ps) marks the maximum safe frequency determined with each technique.

5.2.1 BTI impact on maximum frequency

Table 3 shows the maximum V_{th} degradations found after evaluating BTI aging for all the transistors in the ALU16 circuit. In particular, the absolute maximum ΔV_{th} obtained after 10 years of aging is 58.3 mV (13 %) for a PMOS transistor and 51.4 mV (10.1 %) for an NMOS transistor. These values are consistent with the predictions from prior literature [31].

With our flow, we can transform the ΔV_{th} into switching time degradations. Figure 6 shows how the V_{th} degradations affect the slack time of the different benchmark circuits after one execution period (2 s), using STA and glsdta with no aging, and DTA with BTI aging at the start of the device lifetime (i.e., t = 0 s). In general, the graphs show that using STA alone would limit severely the maximum operating frequency considered safe: in the case of the complete ALU16, the maximum frequency determined with STA is almost a 40 % lower than with DTA with aging. Furthermore, the comparison between the curve for Adder16 and the rest shows also that—at least for small circuits—DTA can neither be easily used on its own because the maximum frequency determined with this technique alone, although generally conservative, may, depending on the concrete characteristics of the workload and circuit under test, move the system into unsafe working conditions. A plausible explanation for this observation is that small circuits can work at higher frequencies; hence, small variations in V_{th} have a bigger absolute impact on frequency than in the case of big circuits, which already require longer periods.

A second observation is that the maximum operating frequency determined for MAC8 is higher than for Mult16, despite the MAC operation being apparently more complex. The reason is that MAC8 performs an $8 \times 8 \rightarrow 16$ bit multiplication, and thus features shorter critical paths than Mult16 (which performs a $16 \times 16 \rightarrow 32$ bit multiplication). The lighter workload of MAC8 also produces lower BTI aging. In the case of ALU16, this effect is masked because both operators share part of their structure and hence are partially subject to similar degradations.

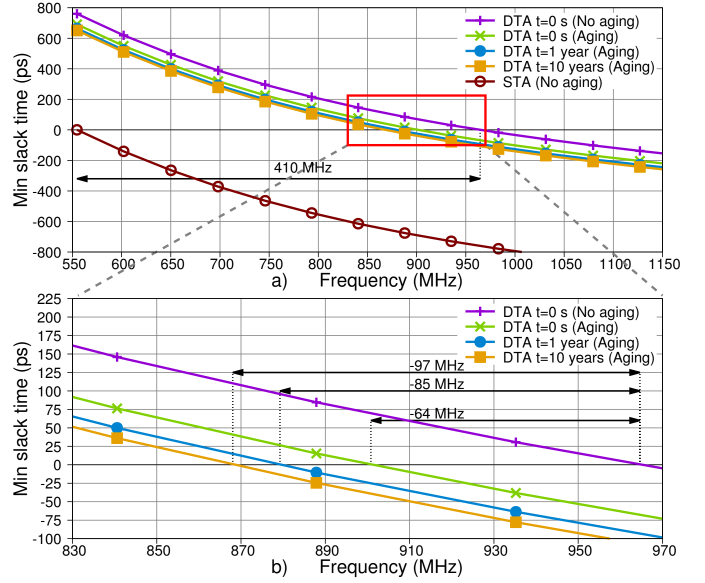


Fig. 7. Minimum slack time at different working frequencies for the ALU16. The X axis (slack time = 0 ps) marks the maximum safe frequency determined with each timing analysis technique and aging time. a) Graph over the full frequency span. b) Zoom over the region where slack time = 0 ps for the DTA analyses.

Figure 7 provides a closer look into the BTI-induced degradations that affect the complete ALU16 circuit after 2 seconds of circuit operation with increasing accumulated aging: 0 seconds, 1 year and 10 years. As expected, most of the timing degradation happens during the first months of the circuit's operating life, corresponding to the filling of the short and mid-term traps. This progressive saturation leads to a reduction of 85 MHz in operating frequency (8.8 %) after 1 year, and up to 97 MHz of reduction (10 %) after 10 years, compared to the maximum frequency obtained with DTA without BTI aging (964 MHz).

These results show that, in the considered experiments, BTI aging has an impact of approximately 10 % ($\sim$100 MHz) on the maximum operating frequency. These results reflect the fact that with the considered biosignal processing workload, the processor spends long periods in sleep mode. Since most of the transistors have enough time to recover (releasing trapped charges) from the stress accumulated during the active periods, only the pseudo-permanent part of the BTI aging is conserved between execution periods. Nevertheless, this performance degradation must be taken into account during the design of the biomedical device, particularly because the impact of BTI is expected to increase with future reductions in transistor size [3], [32]. Alternatively, the tighter determination of safe operating conditions of our framework with respect to STA can enable the selection of lower supply voltages for the same latency target, with the corresponding savings in energy consumption.

5.2.2 BTI impact at the functional level

Our framework can identify the operations that produce erroneous results through a bit-level analysis on the output signals, i.e. by

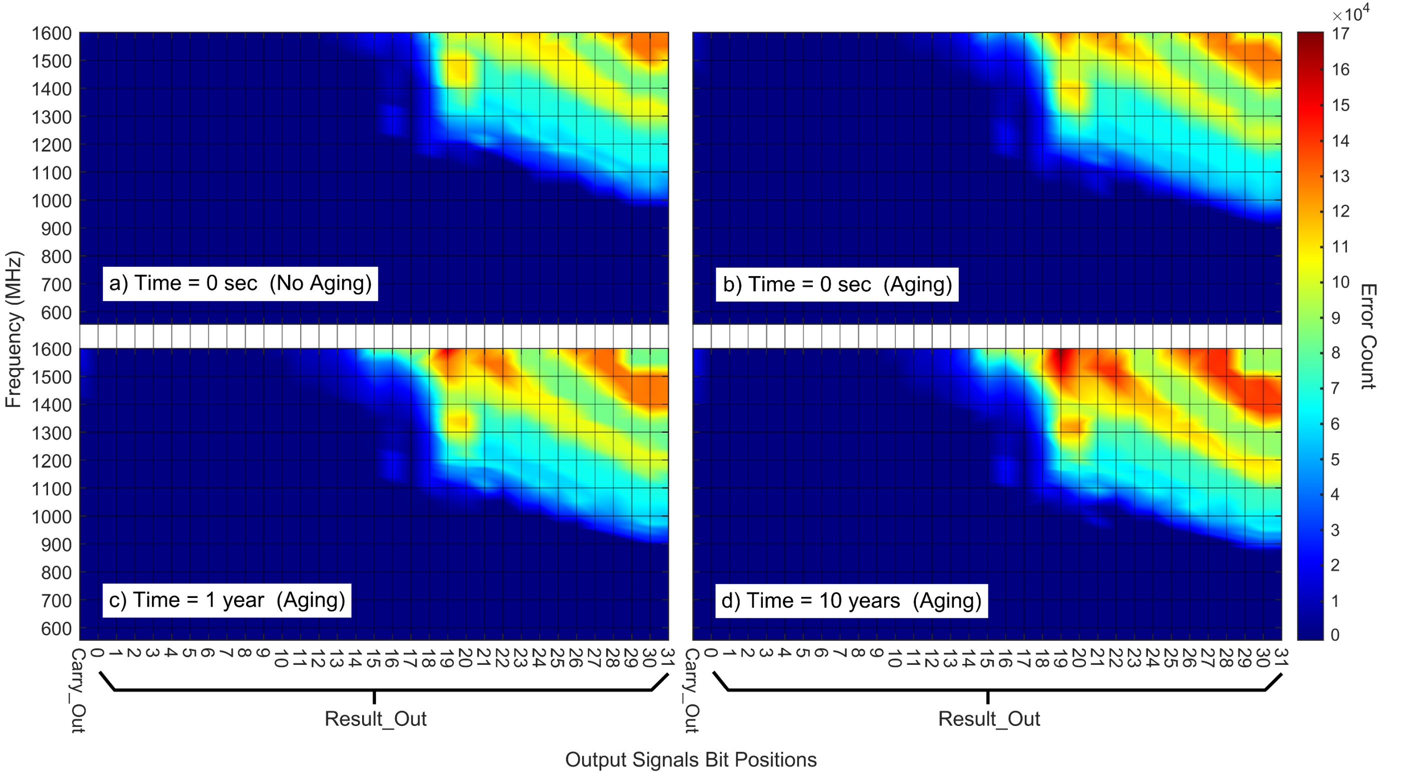
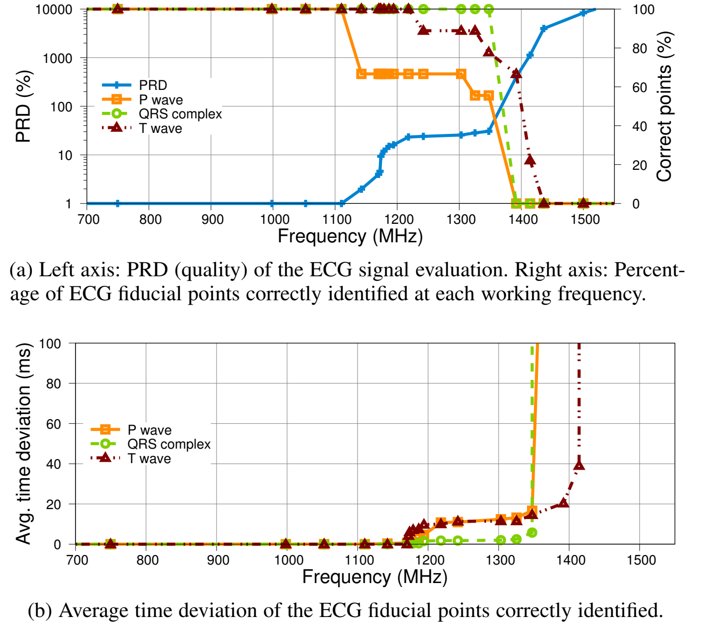


Fig. 8. Heat maps: degradation of the ALU16 output signals (Carry_Out, Result_Out) when increasing the operating frequency, measured after 2 seconds of circuit operation with varying accumulated BTI aging. Aging causes the first erroneous bits to appear at lower frequencies.

using DTA with and without BTI aging. In this way, we can evaluate how timing violations affect the quality of the results delivered by the application and their medical significance. Figure 8 shows how errors accumulate on the output signals of the ALU16 with increasing operating frequencies, during 2 seconds of circuit operation without and with BTI aging (0 seconds, 1 year and 10 years). The important area of the heat maps is the frontier where the first erroneous bit appears: Once a bit is erroneous, the bits with higher weights should be considered as undefined. Therefore, for 10 years of aging, the first errors start to appear (in $\mathrm{Result_{out}}(31)$) at $f \approx 870\,\mathrm{MHz}$. Interestingly, at least for this specific circuit, aging does not change drastically the general shape of the graphs, simply offsetting their features. Similar results are observed for other operations in the ALU16 circuit.

Notwithstanding the previous results, some applications exhibit an inherent resilience against computation errors. Additionally, each timing violation does not necessarily imply an error in the application output, since it may affect signals that are unused in the current combination of operations. Thus, knowing the characteristics of the concrete applications executed on the platform may help to establish more favorable working conditions for the system with the goal of achieving higher frequency or reducing energy consumption. For example, in 3L-MMD, only the 16 LSBs of the results produced by the multiplier are used. Therefore, the application results degrade only when the adder $\mathrm{Carry_{out}}$ or the 16 LSBs of the multiplier output ($\mathrm{Result_{out}}$) are affected by functional errors, which start to happen at $f \approx 1100\,\mathrm{MHz}$ (for $\mathrm{Result_{out}}(15)$). Figure 9 shows that the quality of the delineation is indeed unaffected up to that frequency. Moreover, many high-level biomedical applications, such as epilepsy and obstructive apnea detection, require only the accurate delineation



(a) Left axis: PRD (quality) of the ECG signal evaluation. Right axis: Percentage of ECG fiducial points correctly identified at each working frequency.

(b) Average time deviation of the ECG fiducial points correctly identified.

Fig. 9. Quality assessment of the output generated by the MMD delineation application after 10 years of BTI aging.

of the QRS complex from each heartbeat. In those cases, the system can operate with a $\mathrm{PRD} < 9\,\%$ at a frequency of up to $1172\,\mathrm{MHz}$ ($112\,\%$ over the frequency determined with STA), possibly enabling longer power-gated periods. This possibility is uncovered by the long-term, circuit-level and workload-aware dynamic timing analysis of our flow.

5.2.3 Flow execution time

The most computationally expensive steps of the flow are the switch-level simulations for workload propagation (step 4 in Fig. 2), the ΔV_{th} computation with the BTI model (step 5) and the transistor-level SPICE simulations for DTA (step 7). Producing the results for the ALU16 circuit required in the order of 150 hours in our multicore severs. The proposed long-term BTI evaluation methodology achieved significant time savings. In particular, for just one single transistor, the 1 year ΔV_{th} evaluation based on the exhaustive BTI model requires approximately 5.3 days, while with our proposed methodology the long-term BTI evaluation of all the 16 k transistors that compose the ALU16 takes only 7.4 hours. This represents a reduction in the execution time of the BTI model of four orders of magnitude, proving that our methodology could be applied to complex multiprocessor netlists while keeping acceptable simulation time.

6 Conclusions

In this paper we have presented a complete workload-dependent BTI-aware analysis flow to identify and quantify functional errors on digital synchronous processor architectures along the complete device lifetime. Our flow introduces long-term extrapolation to account for the previously accumulated aging while considering the characteristics of each transistor but, in contrast with previous long-term analysis methods, it resumes a fully detailed BTI modeling to calculate the effect of aging under a real workload and circuit structure at the desired time. Thus, the dramatic reduction in simulation time achieved by our approach, which still keeps an accurate workload representation that enables analysis of instantaneous BTI effects, fully solves the accuracy-execution time trade-off problem, opening the door to the analysis of full multiprocessor netlists.

Our experiments with the execution stage of a processor pipeline expose a variation of up to 54.6 mV (12.1 %) in the threshold voltage of the circuit transistors after one year of continuous operation, with an impact of 8.8 % on the maximum safe operating frequency. Additionally, to fully capture the partially-compensating interactions of the V_{th} shifts that BTI induces on NMOS and PMOS transistors, we have also explored the effects of BTI on complete processor data path level circuits. These experiments show that, whereas STA generally results in overly pessimistic maximum frequency determination, DTA alone can lead to unsafe working points if long-term circuit aging is not taken into account.

Our framework builds the link between BTI-induced V_{th} variations, timing failures and functional errors, identifying the concrete processor operations and input operands that will likely produce those errors. In this way, we have shown how a BTI model that works at the level of individual NMOS or PMOS transistors can be exploited at increasing levels of abstraction. First, it can be used at the circuit level (identifying the interactions along transistor chains). Second, it can be used at the functional unit level (exploring the relationship between maximum operating frequency and the correctness of operator results). Third, it can also be used at the system level (impact on the quality of the delivered results and their medical significance).

We have applied this new technique to the domain of biosignal processing applications for WBSNs because their pseudo-periodic nature interacts with the partially recoverable nature of BTI. The knowledge obtained can be used to steer a graceful degradation of WBSN functionality, or to provide reliability guarantees by reducing frequency or increasing energy consumption if necessary. Alternatively, a careful study of the quality degradation in the application output that can be tolerated versus the expected error rate at a given working frequency enables the design of approximate computing devices that trade off exactness with energy efficiency.

Acknowledgments

This work has been partially supported by the ERC Consolidator Grant COMPUSAPIEN (GA No. 725657), by the EC H2020 WiPLASH (GA No. 863337), and in part by a joint research grant for ESL-EPFL by IMEC.

References

[1] R. Entner, "Modeling and simulation of negative bias temperature instability," Ph.D. dissertation, Technische Universität Wien, Institut für Mikroelektronik, Apr. 2007.

[2] R. Reis, Y. Cao, and G. Wirth, Eds., *Circuit design for reliability*, 1st ed. Springer-Verlag New York, 2015, ch. Recent trends in bias temperature instability, p. 272.

[3] H. Kükner *et al.*, "Scaling of BTI reliability in presence of time-zero variability," in *IEEE Int. Reliability Physics Symp.*, Jun. 2014, pp. CA.5.1–CA.5.7.

[4] D. Stamoulis *et al.*, "Capturing true workload dependency of BTI-induced degradation in CPU components," in *Proc. of the Great Lakes Symp. on VLSI (GLSVLSI)*. ACM, May 2016, pp. 373–376.

[5] P. Weckx *et al.*, "Defect-based methodology for workload-dependent circuit lifetime projections - Application to SRAM," in *IEEE Int. Reliability Physics Symp. (IRPS)*, Apr. 2013, pp. 3A.4.1–3A.4.7.

[6] S. Ramey *et al.*, "Frequency and recovery effects in high-k BTI degradation," in *IEEE Int. Reliability Physics Symp.*, Apr. 2009, pp. 1023–1027.

[7] T. Grasser *et al.*, "The paradigm shift in understanding the bias temperature instability: From reaction–diffusion to switching oxide traps," *IEEE Trans. on Electron Devices*, vol. 58, no. 11, pp. 3652–3666, Nov. 2011.

[8] B. Kaczer *et al.*, "The defect-centric perspective of device and circuit reliability—from individual defects to circuits," in *European Solid State Device Research Conference (ESSDERC)*, Sep. 2015, pp. 218–225.

[9] J. H. Stathis, "The physics of NBTI: What do we really know?" in *IEEE Int. Reliability Physics Symp. (IRPS)*, Mar. 2018, pp. 2A.1–1–2A.1–4.

[10] A. Chaudhary *et al.*, "Consistency of the two component composite modeling framework for NBTI in large and small area p-MOSFETs," *IEEE Trans. on Electron Devices*, vol. 64, no. 1, pp. 256–263, Jan. 2017.

[11] N. Parihar *et al.*, "BTI analysis tool—modeling of NBTI DC, AC stress and recovery time kinetics, nitrogen impact, and EOL estimation," *IEEE Trans. on Electron Devices*, vol. 65, no. 2, pp. 392–403, Feb. 2018.

[12] H. Amrouch *et al.*, "Reliability-aware design to suppress aging," in *ACM/IEEE Design Automation Conference (DAC)*, Jun. 2016, pp. 1–6.

[13] A. Subirats *et al.*, "A new gate pattern measurement for evaluating the BTI degradation in circuit conditions," in *IEEE Int. Reliability Physics Symp.*, Jun. 2014, pp. 5D.1.1–5D.1.5.

[14] V. M. van Santen *et al.*, "Designing guardbands for instantaneous aging effects," in *ACM/IEEE Design Automation Conference (DAC)*, Jun. 2016.

[15] D. Rodopoulos *et al.*, "Atomistic pseudo-transient BTI simulation with inherent workload memory," *IEEE Trans. on Device and Materials Reliability*, vol. 14, no. 2, pp. 704–714, Jun. 2014.

[16] R. Reis, Y. Cao, and G. Wirth, Eds., *Circuit design for reliability*, 1st ed. Springer-Verlag New York, 2015, ch. Compact modeling of BTI for circuit reliability analysis, p. 272.

[17] J. B. Velamala *et al.*, "Logarithmic modeling of BTI under dynamic circuit operation: Static, dynamic and long-term prediction," in *IEEE Int. Reliability Physics Symp. (IRPS)*, Apr. 2013, pp. CM.3.1–CM.3.5.

[18] D. Rodopoulos *et al.*, "Understanding timing impact of BTI/RTN with massively threaded atomistic transient simulations," in *IEEE Int. Conference on IC Design Technology (ICICDT)*, May 2014, pp. 1–4.

[19] Synopsys, "NanoTime STA."

[20] C.-C. Chen *et al.*, "System-level modeling of microprocessor reliability degradation due to BTI and HCI," in *IEEE Int. Reliability Physics Symp.*, Jun. 2014, pp. CA.8.1–CA.8.9.

[21] J. Constantin *et al.*, "Exploiting dynamic timing margins in microprocessors for frequency-over-scaling with instruction-based clock adjustment," in *Design, Automation Test in Europe (DATE)*, Mar. 2015, pp. 381–386.

[22] J. Bhasker and R. Chadha, *Static timing analysis for nanometer designs*. Springer US, 2009.
[23] C. R. Lefurgy *et al.*, “Active guardband management in Power7+ to save energy and maintain reliability,” *IEEE Micro*, vol. 33, no. 4, pp. 35–45, Jul. 2013.
[24] Y. Sun, K. Chan, and S. Krishnan, “Characteristic wave detection in ECG signal using morphological transform,” *BMC Cardiovascular Disorders*, vol. 5, no. 28, Sep. 2005.
[25] F. Rincón *et al.*, “Development and evaluation of multilead wavelet-based ECG delineation algorithms for embedded wireless sensor nodes,” *IEEE Trans. on Information Technology in Biomedicine (T-ITB)*, vol. 15, no. 6, pp. 854–863, Nov. 2011.
[26] Y. Sun, K. Chan, and S. Krishnan, “ECG signal conditioning by morphological filtering,” *Computers in Biology and Medicine (CBM)*, vol. 32, no. 6, pp. 465–479, Nov. 2002.
[27] R. Braojos *et al.*, “Hardware/software approach for code synchronization in low-power multi-core sensor nodes,” in *Design, Automation Test in Europe (DATE)*, Mar. 2014, pp. 1–6.
[28] Y. Zigel, A. Cohen, and A. Katz, “The weighted diagnostic distortion (WDD) measure for ECG signal compression,” *IEEE Trans. on Biomedical Engineering*, vol. 47, no. 11, pp. 1422–1430, Nov. 2000.
[29] A. Darwish and A. E. Hassanien, “Wearable and implantable wireless sensor network solutions for healthcare monitoring,” *Sensors*, vol. 11, no. 6, pp. 5561–5595, 2011.
[30] A. Y. Dogan *et al.*, “Low-Power processor architecture exploration for online biomedical signal analysis,” *IET Circuits, Devices Systems*, vol. 6, no. 5, pp. 279–286, Sep. 2012.
[31] E. Maricau and G. Gielen, “Transistor aging-induced degradation of analog circuits: Impact analysis and design guidelines,” in *Proc. of the ESSCIRC*, Sep. 2011, pp. 243–246.
[32] I. Agbo *et al.*, “Integral impact of BTI, PVT variation, and workload on SRAM sense amplifier,” *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, 2017.

Loris Duch received the M.Sc. degree in Engineering, Physics, Electronics and Materials, specialized in digital integrated circuit design from the Grenoble Institute of Technology (Grenoble INP), France, in 2014. He received a Ph.D. in Microsystems and Microelectronics from EPFL in 2018. His main research interests include hardware and software co-design exploration, energy vs. reliability aware circuit optimization and ultra-low power integrated circuit design, in the field of bio-medical signal processing.

Miguel Peón-Quirós received a Ph.D. on Computer Architecture from UCM, Spain, in 2015. He collaborated as a Marie Curie scholar with IMEC (Leuven, Belgium) and as postdoctoral researcher with IMDEA Networks (Madrid, Spain). He has participated in several H2020 and industrial projects. He is currently a postdoctoral researcher at EPFL. His research includes optimizations for embedded devices and AI for IoT.

Pieter Weckx received the B.Sc. degree in Electronic Engineering, M.Sc. degree in Nanoscience and technology and Ph.D. degree in Engineering from KU Leuven, Belgium, in 2009, 2011 and 2016 respectively. In 2015 he joined IMEC where he is currently a senior researcher working on Design and System technology optimizations for advanced future scaled CMOS. His research interests include HCI, BTI and RTN reliability and their impact on deeply scaled logic technology including SRAM.

Alexandre Levisse received his B.S. and M.S. (Electrical Engineering) degrees in 2014, from Aix-Marseille University, France. In 2017, he received the Ph.D. degree in electrical engineering from CEA-LETI, Grenoble, France and from Aix-Marseille University. He is currently a postdoctoral researcher in the Embedded Systems Laboratory (ESL) of EPFL, Switzerland. His research interests include circuit and architectures for emerging memory and transistor technologies, 3D stacked architectures, in-memory computing and accelerators for neuromorphic computing.

Rubén Braojos is currently a senior engineer and clinical data manager at SmartCardia SA, a bio-medical company in Lausanne, Switzerland. He has been a postdoctoral researcher at the Embedded Systems Laboratory (ESL) of EPFL, where he was involved in collaboration projects between academia and industry and where he drove research in the field of HW/SW co-design of low-power fault-tolerant platforms for bio-signal processing. He received a Ph.D. degree in electrical engineering from EPFL in 2016. He received his B.Sc and M.Sc. degrees in computer science and engineering from UCM, Spain, in 2008 and 2010 respectively.

Prof. Francky Catthoor received the Engineering and Ph.D. degrees in electrical engineering from KU Leuven, Belgium, in 1982 and 1987, respectively. Between 1987 and 2000, he headed several research domains in high level/system synthesis and architectural methodologies, including deep submicron technology and smart photovoltaics, all at the Interuniversity Microelectronics Center (IMEC), Heverlee, Belgium. He is a part-time Full Professor with the Department of Electrical Engineering, KU Leuven. He has been an Associate Editor for IEEE and ACM journals, such as the IEEE Transactions on VLSI Signal Processing and ACM TODAES. He was the Program Chair of several conferences, including ISSS’97 and SIPS’01. He has been elected an IEEE Fellow in 2005.

Prof. David Atienza (M’05-SM’13-F’16) is associate professor of electrical and computer engineering, and heads the Embedded Systems Laboratory (ESL) at EPFL. He received his Ph.D. degree in computer engineering from UCM, Spain, and IMEC, Belgium, in 2005. His research interests include system-level design methodologies for high-performance multi-processor system-on-chip (MPSoC) and low-power IoT systems, including new thermal-aware design for MPSoCs and many-core servers, and ultra-low power edge AI architectures for IoT. He has co-authored more than 300 papers, several book chapters, and eleven patents. He received the DAC Under-40 Innovators Award in 2018, IEEE TCCPS Mid-Career Award in 2018, an ERC Consolidator Grant in 2016, the IEEE CEDA Early Career Award in 2013, and the ACM SIGDA Outstanding New Faculty Award in 2012. He is an IEEE Fellow, an ACM Distinguished Member, and has served as IEEE CEDA President (period 2018–2019).