Robust Processor Based on Alternating Module Activation

Daniel Sarsur Câmara
Department of Electronic Engineering
Federal University of Minas Gerais
Belo Horizonte, Brazil

Lucas Gomes
Department of Electronic Engineering
Federal University of Minas Gerais
Belo Horizonte, Brazil

Frank Sill Torres
Department of Electronic Engineering
Federal University of Minas Gerais
Belo Horizonte, Brazil

Abstract— **The continuously miniaturization of CMOS technology leads to increasing concerns regarding reliability. Thereby, challenges arise from temporary as well as permanent faults. Amongst the various methods proposed in order to diminish the impact of the latter, Alternating Module Activation (AMA) proved to be promising. The aim of this work is the presentation of the implementation of this approach in a proprietary RISC processor. Further, Built-In Self-Test (BIST) capability has been added in order to enable to identification of faulty blocks. Experiments indicate the feasibility of the approach.**

Keywords—Reliability, Robustness, Processor Design, Redundancy, Error detection, BIST.

I. INTRODUCTION

Complementary Metal-Oxide-Semiconductor (CMOS) is still the predominating technology for digital designs with no identifiable concurrence in the near future. Driving forces of this leadership are the high miniaturization capability and the robustness of CMOS. Such technologies, with device dimensions in the range of a few nanometers, suffer from an increased susceptibility to different kinds of failures during operation. In contrast to previous technology generations, solutions within the manufacturing process are not sufficient anymore to deal with these kinds of issues.

Accordingly, reliability concerns are not only an issue of manufacturing anymore, but also have to be considered in all abstraction layers of the design process. Thereby, three main strategies can be identified: (I) design techniques that detect errors, (II) techniques that detect and correct errors, and (III) those techniques that try to avoid or at least prolongate the time until an error.

These concerns are even more significant when related to processors due to its widely use in consumer as well as critical applications. For this reason, researches have been conducted in order to define solutions for either temporary or permanent faults. In case of the former, one can name Roll-Back strategy supported by Bulk-BICS [1-2]. Exemplary techniques related to permanent faults are Redundant Multithreading (RMT) and Chip-level Redundantly Threaded Processor (CRT), for example, and others, based on software solutions.

In this work, we apply a design technique that relates to strategy (III) and combines Sleep Transistors [3] with the idea of modular redundancy to extend the expected lifetime of integrated circuits. This strategy is also combined with techniques of strategy (II) in order to achieve even longer system lifetimes.

The rest of this paper is organized as follows. Section II presents the essential fundamentals to ease the understanding of the following sections while section III summarizes the initial approach and extensions of the design technique. The following section IV presents the strategy to verify the system and, finally, section V present and discusses some results and then, concludes this work.

II. PRELIMINARIES

This section presents basic information regarding the content of this work.

A. Source of Failure

The three main lifetime degrading effects considered in this work are Electromigration (EM), Time Dependent Dielectric Breakdown (TDDB), and Negative Bias Temperature Instability (NBTI). These will be briefly introduced in the following.

In CMOS technologies the best-known failure mechanism is Electromigration which means the transport of material atoms due to the gradual movement of the ions in a conductor caused by the electric current. Due to this atom migration material can be depleted or accumulated. As consequence, high resistive connections or even abrupt breaks can be created. Another result can be undesired connections between interconnects. Equation (1) shows a widely used model for the MTTF due to Electromigration based on Black's Electromigration equation (1) [4]:

$$MTTF_{EM} = A_{EM}(J - J_{crit})^{-n} e^{\frac{E_a}{k_B T}} \qquad (1)$$

with A_{EM} is an empirically determined constant, J is the current density in the interconnect, J_{crit} is the critical current density required for Electromigration, E_a is the activation energy for Electromigration, k_B is Boltzmann's constant, T is absolute temperature in Kelvin, and n is an empirical constant. It can be concluded that during the system's runtime the main driving forces for Electromigration are high currents which lead to high current densities and high temperatures.

Gate oxide breakdown means the formation of a conducting path between the gate and the substrate or source/drain, respectively. The breakdown can be based on abrupt events,

e.g. Electro-Static Discharge, or on destruction over time known as Time Dependent Dielectric Breakdown (TDDB). The latter is due to an autocatalytic loop in which overlapping charge traps create a conducting path between gate and substrate (or source/drain) which leads to increased current flow and heat dissipation. Consequently, thermal damage occurs and more charge traps are created. This positive feedback loop results in an accelerated breakdown and finally in a defect transistor [5]. Following from experimental work, the mean time to failure due to TDDB can be modeled as [6]:

$$MTTF_{TDDB} \propto \left(\frac{1}{V_{DD}}\right)^{a-bT} e^{\frac{X+\frac{Y}{T}+ZT}{kT}} \qquad (2)$$

where V_{DD} denotes the supply voltage, and a, b, X, Y as well as Z are fitting parameters. It can be concluded that at runtime TDDB depends on the applied voltage level at the gate and the temperature.

Negative Bias Temperature Instability (NBTI) is a performance degrading failure mechanism observed mainly in PMOS transistors since they usually operate with negative gate-to-source voltage. This temperature-activated effect occurs when a voltage stress is applied to the transistor gate. The consequence of NBTI is a significantly increase of the transistor threshold voltage V_{th} and following higher delays and leakage currents of the affected integrated system. The physical reasons for NBTI are hole trapping in pre-existent oxide traps and the creation of interface states [7]. Thereby, the interface trap generation $N_{it}(t)$ which leads to a linear increase of V_{th} can be expressed with [8]:

$$N_{it}(t) = 1.16 \sqrt{\frac{k_f N_0}{k_r}} (D_H t)^{\frac{1}{4}} \qquad (3)$$

where the mobile diffusing species are assumed to be neutral H atoms. N_0 is the concentration of initial interface defects, D_H is the corresponding diffusion coefficient, and k_f and k_r are constant dissociation rate and self-annealing rate, respectively. When the device is in recovery phase, k_f becomes zero, and k_r is unchanged. In summary, at runtime NBTI depends on temperature and the applied gate voltage.

B. Power Gating with Sleep Transistors

The application of sleep transistors for power gating is one of the most effective methods to reduce standby leakage [9]. A sleep transistor is referred to either a high threshold voltage PMOS or NMOS transistors which are used as switches to disconnect power supplies from design modules during standby mode. Thereby, the sleep transistors create a virtual power (PMOS) and/or a virtual ground (NMOS). That means in theory, during standby all leakage currents of the gated module are zeroed. It should be noted that even when all sleep transistors are switched off a small leakage component exists. This is mainly based on sub-threshold leakage of the sleep transistors [10].

C. Built-in Self-test (BIST)

A mechanism to identify failure in a part of a device is the Built-in Self-test (BIST). This mechanism consists of test vectors that are executed in order to analyze selected parts of the system. The results of these routines are compared with the expected result. This technique allows the device itself to identify particular faults. With this diagnosis it is possible that the system make decisions to overcome these flaws, or at least, to generate a warning.

D. Alternating Module Activation (AMA)

An essential characteristic of power gating with Sleep Transistors is its ability to dynamically disconnect the power supply during the runtime of integrated systems. This implies that in this case, ideally, there are no inherent currents and voltages, and thus, electromagnetic fields. Furthermore, local temperatures are reduced as there is no switching activity present. Considering all this, it is easy to see that all the lifetime degrading effects quoted above are eliminated or at least strongly reduced. As a consequence, the Mean Time To Failure (MTTF) is prolonged approximately by the time that the design is in the idle phase.

The initial approach Alternating Module Activation (AMA) proposes that each gated module is implemented at least twice. During the runtime, though, only one of these instances is active while the others are disconnected from the power supply. In order to properly work, additional logic is required to multiplex the results from the currently active instance to the subsequent module as it is illustrated in Fig. 1 (grayed out parts).

E. Enhanced AMA (EAMA)

The improved approach Enhanced Alternating Module Activation (EAMA) proposes extensions to the AMA approach by increasing the lifetime in case of faultiness of one of the instances. Beside this, error detection capability and Built-in self-test (BIST) functionality are added (see Fig. 1). Just as AMA, a control logic (omitted in Fig. 1) controls which of the instances is active and which is disconnected. Further, comparators are added to the multiplexers mentioned above to verify that every multiplexer's input has the same value during the transition phase in which both instances are active.

If a different result is detected, the logic block suspends the program and activates a BIST. A test program is executed consecutively in both instances. The results are compared with an expected value and, in case of detection of an error the faulty instance is suspended and only the other instance remains connected.
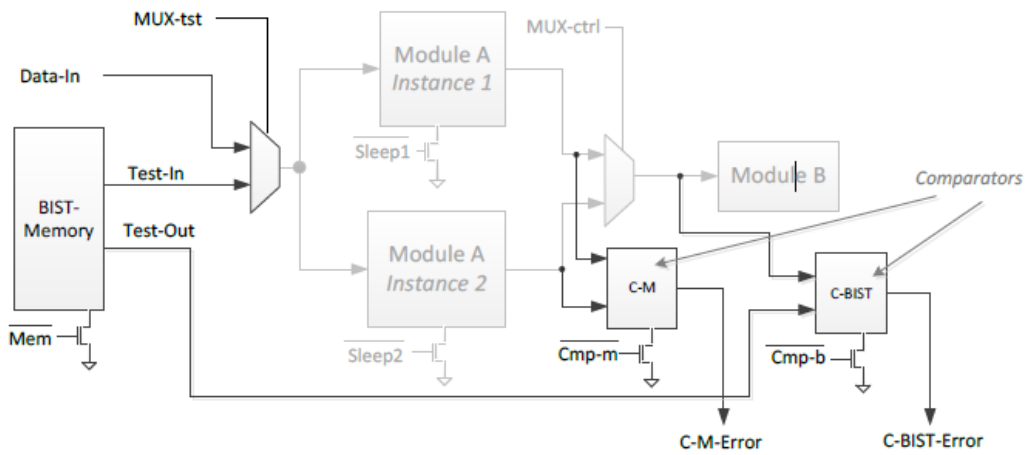
Fig. 1. Structure of Enhanced AMA. Scheme with two redundant instances, comparators for error detection and BIST functionality

III. IMPLEMENTATION

In order to test the feasibility of this approach we first developed a pipelined RISC processor. It is similar to the MIPS architecture and consists of a five-pipeline-stages structure. Further, a cache memory control circuit had been added.
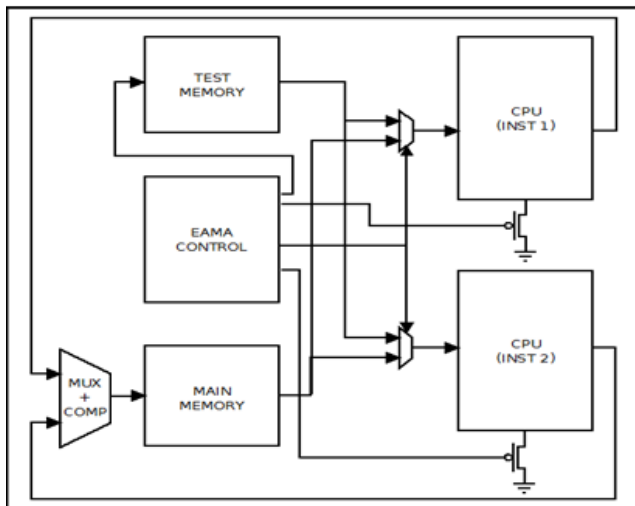

Fig. 2. No. of logical elements for each kind of implementation

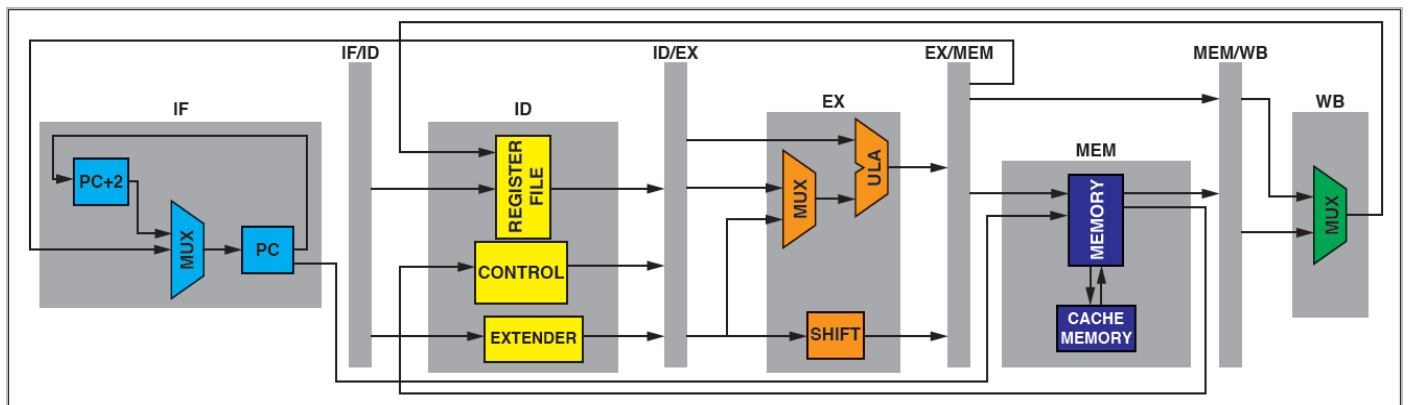Fig. 2 presents the implemented stages, namely Instruction Fetch (IF), Instruction Decode (ID), Execute (EX), Memory Access (MEM) and Write Back (WB). Furthermore, the main components of each stage and the registers, which store data and signals between stages, were added.

In the next step, both techniques AMA and EAMA had been implemented. Initially, the CPU pipeline was duplicated. This was followed by the implementation of the multiplexers required to control the data flow and the logic block controlling the activation and deactivation of the instances.

Subsequently, we implemented the EAMA technique by creating comparators that would be able to compare some results of both instances during the transition phase when both are active. As it is only possible to compare the results during a limited period of time, it is not intended to identify transient faults, but to detect permanent faults. A simplified schematic can be seen in Fig. 3.

In case of detection of any error, *i.e.* different values in same outputs of the instances, the logic control block stops the program execution and activates a BIST. In order to test the greatest amount of blocks in one instance we created a program that applies almost all instructions and is stored in another memory. When BIST is active, it turns the first instance on, runs the test program and verifies its result. Then it turns the first instance off, turns the second instance on, runs the test program again and verifies the second result. As soon as it has tested both instances the logic block can decide whether to suspend one or another instance operation. A flowchart can be seen in Fig. 4.
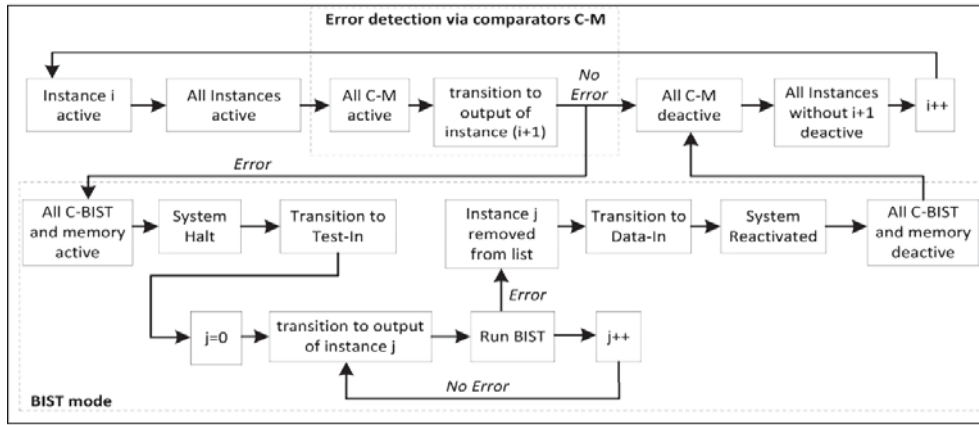

Fig. 3. CPU pipeline overview

Fig. 4. Control flow for extended AMA approach, enhanced by an error detection phase during module transition and a BIST mode

IV. VERIFICATION AND RESULTS

In order to verify the correct operation of the whole system, we created a new block containing the processor and a logic unit that controls whether there will be an error or not and where the error will occur. For error insertion, we placed several multiplexers in neuralgic points at which we were able to change a signal of normal flow by the user.

These tests were performed in an Altera DE2-70 Development Board that contains a Cyclone II FPGA (model EP2C70F896C6). This board also allows the selection of different input data by using switches, and to show information related to outputs and errors by using LEDs and the LDC display. Fig. 5 depicts the switches, keys and leds used on the board and their function.

The experiments indicated that the system operates as described above, being able to detect errors in different parts of the circuit and suspend the operation of the instance where the error was found.

Fig. 5 depicts the area of each implementation in terms of Logic Elements (LE) occupied on the FPGA. It follows that the AMA implementation increased the area by 15.7 %, while the EAMA approach required 23.9 % more LE, both compared to the initial CPU pipeline. It should be noted that all memory elements are implemented with LE. The maximum frequency decreased from 67.1 MHz of the raw CPU pipeline to 62.6 MHz (AMA) and 46.8 MHz (EAMA).
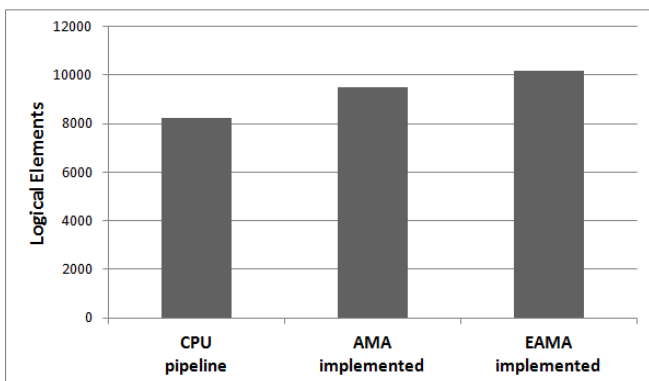

Fig. 5. No. of logical elements for each kind of implementation

V. RESULTS AND CONCLUSION

This work presents the implementation of the approach Alternating Module Activation (AMA) in a RISC processor in order to extend the expected lifetime. It could be shown that although its complexity, it is possible to integrate the (E)AMA technique either in simple devices as in a robust processor. Naturally, additional logic is required and there is an inevitable trade-off between many aspects such as area, power consumption, system lifetime and Mean Time To Failure (MTTF).

In future steps we will continue with the digital project flow including synthesis for an 180nm CMOS process, floorplaning, placement and routing with the ultimate goal of assembling a chip. We will also develop an environment to analyze aging effects.

REFERENCES

[1] R. K. V. Maeda, F. Sill T., "RISC Processor with Single Event Transient Detection and Instruction Roll-Back", SBMicro, Sforum, 2013

[2] R. P. Bastos, F. Sill T., "Detection of Transient Faults in Nanometer Technologies by using Modular Built-In Current Sensors", Minas gerais, Brazil

[3] Cornelius, Claas ; SILL , F. ; Timmermann, Dirk. Power-Efficient Application of Sleep Transistors to Enhance the Reliability of Integrated Circuits. Journal of Low Power Electronics (Print), v. 7, p. 552-561, 2011.

[4] J.R.Black, "A brief survey of electromigration and some recent results", In IEEE Transactions on Electron Devices, 1969.

[5] D. Crook, "Method of determining reliability screens for time dependent reliability breakdown", IRPS, 1979.

[6] E. Y. Wu et al., "Interplay of voltage and temperature acceleration of oxide breakdown for ultra-thin gate dioxides", Solid-state Electronics Journal, 2002.

[7] T. Grasser and B. Kaczer, "Evidence that two tightly coupled mechanisms are responsible for negative bias instability in oxynitride MOSFETs", IEEE Trans. Ele.. Devices, 2009, 56, (5), pp. 1056–1062.

[8] E. Maricau and G. Gielen, "NBTI model for analogue IC reliability simulation", Electronics Letters, 2010, 46.

[9] A. Ramalingam, B. Zhang, A. Davgan, and D. Pan, "Sleep transistor sizing using timing criticality and temporal currents", Proc. ASP-DAC, 2005.

[10] M. Powell, S.-H Yang, B. Falsafi, K. Roy, and T. N. Vijaykumar, "Gated-Vdd: A circuit technique to reduce leakage in deep-submicron cache memories", Proc. ISLPED'00, 2000, pp. 90-95.