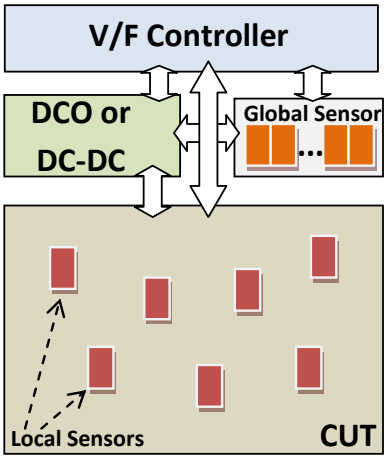
UAlg
UNIVERSIDADE DO ALGARVE

UNIVERSIDADE DO ALGARVE
INSTITUTO SUPERIOR DE ENGENHARIA



DYNAMIC POWER AND FREQUENCY OPTIMIZATION IN DIGITAL ELECTRONIC SYSTEMS

OPTIMIZAÇÃO DINÂMICA DA TENSÃO DE ALIMENTAÇÃO E DA FREQUÊNCIA DE OPERAÇÃO EM SISTEMAS ELECTRÓNICOS DIGITAIS

André Azevedo de Sousa Romão

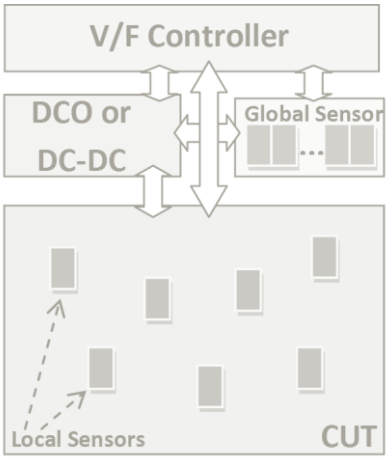
Dissertation for obtaining the Master degree in
Electrical and Electronic Engineering
Specialization Area in Information Technologies and Telecommunications

Tutor: Professor Doutor Jorge Filipe Leal Costa Semião

September, 2013

UAlg
UNIVERSIDADE DO ALGARVE

UNIVERSIDADE DO ALGARVE
INSTITUTO SUPERIOR DE ENGENHARIA



DYNAMIC POWER AND FREQUENCY OPTIMIZATION IN DIGITAL ELECTRONIC SYSTEMS

OPTIMIZAÇÃO DINÂMICA DA TENSÃO DE ALIMENTAÇÃO E DA FREQUÊNCIA DE OPERAÇÃO EM SISTEMAS ELECTRÓNICOS DIGITAIS

André Azevedo de Sousa Romão

Dissertation for obtaining the Master degree in
Electrical and Electronic Engineering
Specialization Area in Information Technologies and Telecommunications

Tutor: Professor Doutor Jorge Filipe Leal Costa Semião

September, 2013

Title: Dynamic Power and Frequency Optimization in Digital electronic Systems.

Authorship: André Azevedo de Sousa Romão.
I hereby declare to be the author of this original and unique work. Authors and references in use are properly cited in the text and are all listed in the reference section.

_____

André Azevedo de Sousa Romão

To my son David Romão
and to my wife Raquel Romão,
with love.

ACKNOWLEDGMENTS

First of all, I would like to thank to Dr. Jorge Semião, my tutor, for his personal and professional dedication, support and guidance in the elaboration of this dissertation. Without his help I would be completely lost, because the research area was new to me. He shared his knowledge, expertise and time, and that allowed me to finish this task.

To my wife Raquel Romão, who always believed in me and for all the help she provided. I love you and our son David Romão.

To my father Joaquim Romão and my mother Teodolinda Azevedo, that gave me good advices, on my personal, academic and professional and encouraged me to pursue this goal.

I also like to thanks to all my friends for the good break times that allow me to refill the energies and to fulfill this thesis.

Thank you all.

André Romão, Faro, September 30, 2013

ABSTRACT

The work developed consists in a power or frequency optimization methodology, for long-term operation, using global and local performance sensors. The methodology allows circuits to be optimized dynamically, during their life-time, according with one of the two possible needs: (1) restrict power consumption, by reducing power-supply voltage to the minimum value that prevents errors from happening; or (2) optimize performance, by increasing operating frequency to the maximum limit that prevents errors' occurrence.

The use of both global and local sensors allows an on-line performance monitoring with information from the critical elements in the circuit and being conscious with the dynamic evolution of different variables that affect performance (e.g., aging degradation, temperature and power-supply variations).

Global sensors are used to perform periodic on-line monitoring. Local sensors are used to keep global sensors tuned with circuit's performance. Both are sensitive to PVTA (Process, power supply Voltage, Temperature and Aging) variations, or to any other parameter that may affect performance, and both can trigger the supply Voltage or Frequency on-line tuning, along product lifetime, preventing functional errors.

Global sensor provides a coarse performance evaluation, while local sensors provide a fine grain performance evaluation on the real functional circuit.

Additionally, a new Delay-Fault Correction Flip-flop (DFC-FF) was presented, to improve the existing local aging sensors, and to allow not only the prediction of errors but also their correction, under certain limits.

Methodology was demonstrated with HSPICE simulations.

KEYWORDS: Performance sensor, Dynamic Voltage Frequency Scaling (DVFS), Frequency and Power optimization, Aging prediction, Error-Correction.

RESUMO

À medida que a tecnologia de circuitos integrados CMOS é exposta à miniaturização, surgem diversos problemas no que diz respeito à fiabilidade e performance. Efeitos tais como o BTI (Bias Thermal Instability), TDDB (Time Dependent Dielectric Breakdown), HCI (Hot Carrier Injection), EM (Electromigration) degradam os parâmetros físicos dos transístores CMOS e por sua vez alteram as propriedades eléctricas dos mesmos ao longo do tempo. Esta deterioração é chamada de envelhecimento e estes efeitos são cumulativos e têm um grande impacto na performance do circuito, especialmente se ocorrerem outras variações paramétricas, como as variações de processo, temperatura e tensão de alimentação. Estas variações são conhecidas por variações PVTA (variações no Processo de Fabricação do circuito integrado [P], na Tensão de Alimentação [V], na Temperatura [T] e variações provocadas pelo Envelhecimento dos circuitos [A]) e podem desencadear erros de sincronismo durante a vida do produto (circuito integrado digital).

O trabalho apresentado nesta dissertação tem por objectivo primordial o desenvolvimento de um sistema que optimize a operação ao longo da vida de circuitos integrados digitais síncronos de forma dinâmica. Este sistema permite que os circuitos sejam optimizados de acordo com as suas necessidades:

(i) Diminuir a dissipação de potência, por reduzir a tensão de alimentação para o valor mais baixo que garante a operação sem erros; ou

(ii) Aumentar o desempenho/performance, por aumentar a frequência de operação até ao limite máximo no qual não ocorrem erros.

A optimização dinâmica da operação ao longo da vida de circuitos integrados digitais síncronos é alcançada através de um controlador, um bloco de sensores globais e por vários sensores locais localizados em determinados flip-flops do circuito.

A nova solução tem como objectivo utilizar os dois tipos de sensores atrás mencionados, globais e locais, para possibilitar a previsão de erros de performance de uma forma mais eficaz, que possibilite a activação de mecanismos que impeçam a ocorrência de erros durante o tempo de vida útil de um circuito, e dessa forma

permitindo optimizar constantemente o seu funcionamento. Assim é exequível desenvolver circuitos que operem no limite das suas capacidades temporais, sem falhas, e com a utilização de margens de erro pequenas para admitir as variações de performance provocadas por variações no processo de fabrico, na tensão de alimentação, na temperatura ou o envelhecimento.

Foi também desenvolvido um sistema de controlo que permite, depois da detecção de um potencial erro, desencadear um processo para diminuir a frequência do sinal de relógio do sistema, ou aumentar a tensão de alimentação, evitando que o erro ocorra.

Apesar de existirem outras técnicas de controlo dinâmico da operação de circuitos integrados tais como DVS (Dynamic Voltage Scaling), de DFS (Dynamic Frequency Scaling), ou ambas (DVFS – Dynamic Voltage and Frequency Scaling), estas técnicas ou são de muito complexa implementação, ou apresentam margens de segurança elevadas, levando a soluções em que a operação do circuito não está optimizada. A solução desenvolvida neste trabalho, em que se utilizam sensores preditivos locais e globais os quais são sensíveis ao envelhecimento a longo prazo ocorrido nos circuitos, constitui uma novidade no estado da técnica relativamente ao controlo de sistemas de DVS e/ou DFS.

Outro aspecto importante é que neste trabalho desenvolveu-se um método de ajuste da tensão de alimentação ou da frequência, o qual é sensível ao envelhecimento a longo prazo dos circuitos, utilizando sensores locais e globais. O controlador permite a optimização da performance dos circuitos através do aumento da frequência de operação até ao limite máximo que ainda evita a ocorrência de erros e a optimização de consumo de energia através da redução da tensão de alimentação (VDD) para o valor mínimo que ainda previne a ocorrência de erros.

Através de uma análise de previsão de envelhecimento, são identificados os caminhos críticos, bem como os caminhos que envelhecem mais rápido e que se tornarão críticos com o envelhecimento do circuito. Uma vez identificados os caminhos críticos, irão ser inserido os sensores locais através da substituição dos flip-flops que terminam os caminhos críticos identificados por novos flip-flops que incluem sensores de performance e de envelhecimento. É de referenciar que estes sensores são preditivos, ou seja, que sinalizam precocemente os erros de performance, antes de eles ocorrerem nos flip-flops que capturam os caminhos críticos. A arquitectura dos sensores propostos é tal que as variações PVTA que ocorrem sobre

eles fazem aumentar a sua capacidade de prever o erro, ou seja, os sensores vão-se adaptando ao longo da sua vida útil para aumentarem a sua sensibilidade.

Os sensores locais têm como função realizar a calibração dos sensores globais, bem como realizar a monitorização constante dos atrasos nos caminhos mais longos do circuito, sempre que estes são activados.

A função dos sensores globais é a realização da monitorização periódica ou quando solicitado dos atrasos no circuito digital. Ambos os tipos de sensores, os sensores globais como os locais podem desencadear ajustes na frequência ou na tensão de alimentação.

Os sensores globais são compostos por uma unidade de controlo do sensor global, que recebe ordens do controlador do sistema para iniciar a análise ao desempenho do circuito e gera os sinais de controlo para a operação de análise global do desempenho e por duas cadeias de portas (uma com portas NOR e outra com portas NAND), com tempos de propagação superiores aos caminhos críticos que se esperam vir a ter no circuito durante a sua vida útil. Ambos os caminhos irão, presumivelmente, envelhecer mais que os caminhos críticos do circuito quando sujeitos ao efeito BTI (que influencia fortemente a degradação do Vth dos transístores [NBTI/NORs e PBTI/NANDs]). Ao longo das duas cadeias, diversos sinais à saída de algumas portas NOR e NAND são ligados a células de sensores globais, criando diversos caminhos fictícios com diferentes tempos de propagação. As saídas dos sensores das duas cadeias formam duas saídas de dados do sensor global.

A fim de se alcançar a optimização do desempenho do circuito, são realizados testes de calibração dos sensores, onde são estimulados alguns caminhos críticos no circuito (através de um teste determinístico) e, simultaneamente é realizada a análise do desempenho pela unidade de sensores globais. Este procedimento, permite definir o limite máximo (mínimo) para frequência (tensão de alimentação) sem que os sensores locais sejam sinalizados. Esta informação da frequência (tensão) é guardada num registo do controlador (registo V/F) e corresponde à frequência (tensão) normal de funcionamento. Este teste também permite determinar quais os caminhos fictícios nas duas cadeias que apresentam tempos de propagação semelhantes aos caminhos críticos do circuito. Esta informação também é guardada em dois registos no controlador do sistema (registos GSOsafe), que indicam o estado das saídas dos controladores globais para a operação optimizada do circuito.

Durante a vida útil do circuito, o controlador do sistema de optimização procede ao ajuste automático da frequência (ou da tensão de alimentação) do circuito, caso o controlador dos sensores globais detecte uma alteração em relação à operação correcta em memória, alterando o conteúdo do registo que guarda a frequência (tensão) de trabalho.

Se por ventura ocorrer a sinalização de um sensor local e não existir nenhuma sinalização para alteração do desempenho pelos sensores globais, quer dizer que o circuito pode ter envelhecido mais que os caminhos fictícios dos sensores globais, pelo que a frequência (tensão de alimentação) de funcionamento deve ser alterada, mas também deve existir uma actualização nos registos que guardam a saída correcta dos sensores globais.

É de salientar que, se os caminhos fictícios envelhecem mais do que o circuito, as margens de segurança (time slack) existentes vão sendo aumentadas ao longo da vida do circuito, tratando-se de uma segurança positiva. Mas, se existir a possibilidade do envelhecimento ser maior nos caminhos do circuito, a existência dos sensores locais a monitorizar a todo o tempo o desempenho do circuito, garantem que o sistema pode aprender com as sinalizações e adaptar-se às novas condições de operação ao longo da vida útil do circuito.

Enquanto a monitorização efectuada pelo bloco de sensores globais fornece uma avaliação grosseira do estado de funcionamento do circuito, a monitorização efectuada pelos sensores locais, quando activados, fornece uma avaliação fina sobre qual a performance do circuito para que não ocorram erros funcionais.

As novidades apresentadas neste trabalho são no mecanismo de controlo que permite a optimização dinâmica da tensão ou da frequência, e na arquitectura e funcionamento do sensor global a inserir no circuito. No que diz respeito ao mecanismo de controlo do sistema de optimização dinâmica, as novidades são: (i) na utilização conjunta de sensores locais e globais para garantir níveis de optimização elevados, (ii) na utilização de sensores preditivos (globais e locais) que previnem os erros de ocorrerem e (iii) na utilização de sensores sensíveis ao envelhecimento do circuito ao longo da sua vida útil. Em relação ao sensor global para monitorização de variações PVTA a novidade consiste (iv), na apresentação de sensores para a degradação nos transístores PMOS e de sensores para a degradação nos transístores NMOS.

Este método de optimização e as topologias apresentadas podem ser desenvolvidas e utilizadas com outros tipos de flip-flops, ou empregando outros tipos de sensores, ou outros caminhos fictícios nos sensores globais, sem prejuízo do método global de optimização que conjuga os dois tipos de sensores, globais e locais, para optimizar a tensão de alimentação e a frequência de operação.

É proposta uma nova arquitectura para um flip-flop com correcção de erros de atraso (DFC-FF / AEPDFC-FF) com e sem previsão de erros adaptativa para realizar a correcção/monitorização e correcção on-line da perda de performance a longo prazo de sistemas digitais CMOS, independentemente da sua causa. O DFC-FF integra um FF do tipo TG-MSFF (Transmission Gate Master Slave Flip-Flop) e um sensor de correcção de erros (CES) dos quais são apresentados duas propostas. O AEPDFC-FF é composto por DFC-FF e um sensor de envelhecimento.

A variabilidade tornou-se na principal causa de falha dos circuitos digitais quando a tecnologia evoluiu para as escalas nanométricas. As reduzidas dimensões físicas dos novos transístores e o aumento na complexidade dos circuitos integrados tornou os novos circuitos mais susceptíveis a variações no processo de fabrico, nas condições de operação e operacionais, tendo como consequência o fabrico de dispositivos mais frágeis, com maior probabilidade de falharem nos primeiros meses de vida, e com tempos de vida útil esperados inferiores aos das tecnologias anteriores.

Face a outras propostas, uma das principais vantagens do DFC-FF é que a a perda de performance do próprio sensor melhora a sua capacidade de correcção de erros. Os efeitos do envelhecimento, do aumento de temperatura e da diminuição na tensão de alimentação (VTA), aumentam a janela de correcção, permitindo que o DFC-FF possa estar sempre ligado sem comprometer o seu funcionamento.

O conceito, estudado e desenvolvido em tecnologia de 65nm, pode ser transportado posteriormente para nanotecnologias mais recentes, usando MOSFETs de menor dimensão, uma vez que a arquitectura do sensor é transversal a toda a tecnologia CMOS.

PALAVRAS-CHAVE: Sensor de Performance, Escalonamento Dinâmico da Tensão e da Frequência (DVFS), Optimização da Frequência e da Tensão, Predição do Envelhecimento, Correcção de Erros.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AEP-FF	*Adaptive Error-Prediction Flip-Flop* ou Flip-Flop com Predição Adaptativa de Erro
ASIC	*Application Specific Integrated Circuit*, ou circuito integrado de aplicação específica
ATPG	*Automatic Test Pattern Generator*, ou gerador automático para padrões de teste
IC	*Integrated Circuito* ou Circuito Integrado
CMOS	*Complementary Metal-Oxide Semiconductor*, ou semiconductor de metal e óxido de lógica complementar (família lógica)
CP	*Critical Path*, ou caminho crítico (o caminho com maior tempo de propagação)
CST	*Controller and Sensor Tuning*
EM	*Electro-Migration*, ou migração eléctrica
FF	Flip-flop
FPGA	*Field Programmable Gate Array*, ou matriz de portas programáveis
H	Hidrogénio (símbolo químico)
HCA	*Hot Carrier Aging*, ou envelhecimento por portadores quentes
HCI	*Hot Carrier Injection*, ou inserção de portadores quentes
IT	*Interface Traps*, ou obstáculos à passagem da corrente eléctrica, formados no interface Si-SiO_2
LTD	*Long Term Degradation*, ou degradação em períodos de tempo longos (referente a modelo de degradação)
NBTI	*Negative Bias Temperature Instability*, ou instabilidade na temperatura em polarização negativa
NMOS	*N-type Metal-Oxide Semiconductor*, ou semicondutor de metal e óxido do tipo N (transístor ou família lógica)

NMOSFET	*N-type Metal Oxide Semiconductor Field-Effect Transistors*, ou transístor semicondutor de metal e óxido de efeito de campo do tipo N
PBTI	*Positive Bias Temperature Instability*, ou instabilidade na temperatura em polarização positiva
PDF	*Path Delay Faults*, ou faltas de atrasos em caminhos
PI	*Primary Input*, ou entrada primária
PMOS	*P-type Metal-Oxide Semiconductor*, ou semicondutor de metal e óxido do tipo P (transístor ou família lógica)
PMOSFET	*P-type Metal Oxide Semiconductor Field-Effect Transistors*, ou transístor semicondutor de metal e óxido de efeito de campo do tipo P
PO	*Primary Output*, ou saída primária
PTM	*Predictive Technology Model*, ou modelo preditivo da tecnologia (referente a modelos SPICE de transistores)
PVT	*Process, power-supply Voltage and Temperature*, ou processo (V), tensão de alimentação (V) e temperatura (T) (variações)
PVTA	*Process, power-supply Voltage, Temperature and Aging*, ou processo (V), tensão de alimentação (V), temperatura (T) e envelhecimento (A) (variações)
R-D	*Reaction – Diffusion*, ou reacção – difusão (modelo)
SI	*Secondary Input*, ou entrada secundária
Si	Silício (símbolo químico)
SiO2	Dióxido de Silício
SIV	*Stress Induced Voids*, lacunas induzidas por stresse
SM	*Stress Migration*, ou migração devida ao stresse
SO	*Secondary Output*, ou saída secundária
T	*Temperature* or Temperatura
TC	*Thermal Cycling*, ou fase termal
TV	Tabela de Verdade
TDDB	*Time Dependent Dielectric Breakdown*, ou quebra no dieléctrico dependente do tempo
TID	*Total Ionizing Dose*, ou dose de ionização total (efeito cumulativo da ionização de um circuito ou semicondutor)

V	*Power-supply Voltage*, ou tensão de alimentação (variações)
VHDL	*VHSIC (Very High Speed Integrated Circuits) Hardware Description Language*
VT	*Power-supply Voltage and Temperature*, ou tensão de alimentação (V) e temperatura (T) (variações)
Vth	*Threshold Voltage*, ou tensão limiar de condução dos transístores MOSFET (valor que pode variar)

1. INTRODUCTION

Like living organisms, integrated circuits have during its lifetime a degradation process of the initial characteristics very similar to physiological aging. Thus, this degradation process is also called aging. This is a slow and cumulative process that occurs in the circuits caused by operation time and operating conditions. The origin can be several factors, such as temperature, voltage, humidity, operation frequency, and others [1][2][7]. Aging is caused by all the mechanisms that change the circuit physical and electrical parameters reducing, this way, it is useful lifetime [2].

In addition, physiological aging usually causes decreased responsiveness to stimulus, which in digital circuits can be represented by the response time of the cells, or by the propagation time of the combinational paths associated with the lowest average conductivity of the transistors [3]. To deal with propagation time errors a safety time margin is added (usually defined as time slack, which is the time left in a period, after subtracting the propagation time of the path) to guarantee that the increased delays within the period of operation of the circuits, will not lead to a synchronism error [4]. However, if the propagation time increases further, boosted by variations in the manufacturing process (Process variations), or by variations in voltage supply (power-supply voltage variations), or by variations in the operating temperature (Temperature variations), or by aging (Aging variations), or by the sum of all effects (PVTA variations), there is a synchronism error [5]. This happens when memory cells (usually flip-flops) do not capture the correct value, but a value delayed due to the increase of the propagation time of the combinatorial paths [6]. Thus, PVT variations may cause the same errors that timing variations cause by aging. And these aging variations are cumulative, reducing the lifetime of the integrated circuit (IC), because they become slower and with higher probability to fail.

Once again, is interesting the remarkable resemblance between living organisms and circuits. Gradual degradation by aging changes the electrical parameters of the transistors and as a consequence changes the electrical parameters of the circuit [7],

until the point that the equipment as a whole will stop working due to critical failures (equivalent to death in living organisms).

As mentioned, PVTA are key variations that could cause a circuit to fail. In our days, with the demand of limited energy and power consumption for mobile applications (cell phones, laptops, other battery powered devices), it became necessary to use *Low Power* (LP) design techniques for successful product development. *V_{DD} reduction* and *clock frequency reduction* are two techniques to reduce power dissipation in CMOS. However, parametric variations on the voltage (V_{DD}) introduces additional constraints on the fault-free operation of a circuit, and commonly frequency reduction is also used simultaneously, to avoid delay-fault occurrence. But certain product's specifications may not comply with an overall performance loss.

Nowadays there are several low power techniques. Among dynamic power control techniques are: clock gating, multi power supply voltage (multi-V_{DD}), variable frequency, and power efficient circuits [1]. Leakage power control techniques comprise power gating and multi threshold (multi-V_{th}) cells. Depending on the design complexity and requirements combination of LP techniques are adopted.

1.1 PROBLEM ANALYSIS

As mentioned, CMOS circuit's performance is affected by many parametric variations, like Process, power-supply Voltage and Temperature (PVT) [26], as well as to aging effects (PVT and Aging – PVTA), with the Bias Temperature Instability (BTI) being one of the main concerns that affects PMOS (Negative BTI, NBTI) and NMOS (positive BTI, PBTI) transistors. All these variations degrade circuit performance and increase variability in CMOS circuits, mainly in nanometer technologies. This may lead to chip failures, especially when several effects occur simultaneously or when cumulative degradations pile up. To deal with nanometer technologies increased variability in CMOS circuits, designers use conservative approaches, with wider relative circuits' time slacks (error margins), leading to non-optimized circuits.

As technology scales down, power density is increasing, as the areas shrink and the frequencies increase, but power supply voltages stop scaling down significantly below 1 volt [29]. Dynamic power is still the main concern; however from 65nm and below, leakage power may become similar [30]. Standard techniques like clock-gating are being used to reduce dynamic power in recent server-class microprocessors [29]. However, aggressive techniques to reduce power consumption are gaining their way in battery operated applications (like Dynamic Voltage and Frequency Scaling – DVFS [12]).

Various sensor topologies have been presented, to globally detect circuit's performance degradation [27] (some of them identified as aging sensors). However, Global Sensors age differently than the circuit and consequently, either (1) low slack margins are used allowing circuit failure, because the circuit's critical path (CP) has higher delay degradation than the sensor, or (2) large slack margins are used, and the circuit is not optimized. More recently, Local Sensors were also presented to detect performance errors ([23], [4], [5], [16]). However, Local Sensors will only detect abnormal delays if the critical paths they monitor are activated. Therefore, they work well in off-line mode, but they are not reliable for avoiding errors in on-line operation.

1.2 OBJECTIVES

The main objective of this dissertation consists in implementing a complete aging-aware power or frequency optimization methodology, that addresses previous performance sensors' problems, and is a continuation of the work reported in [25]. In [25] both Global Sensors and Local Sensors are used to permit a constant and on-line performance monitoring, and to allow an optimized operation, targeting reduced power consumption or high performance. Performance optimization is achieved by increasing operating frequency to the maximum limit while still preventing errors from happening. Power consumption optimization is achieved by reducing power-supply voltage (VDD) to the minimum value which can still prevent errors occurrence. The use of global and local sensors allow us to take advantage of the best of each sensor type: global sensors perform an on-line effective monitoring, whereas local sensors predict errors locally in the circuit, where they really occur.

However, in [25] the controller to allow the complete Voltage or Frequency optimization was not implemented, it was only defined its simple behavior. Moreover, a digitally-controlled VCO (Voltage Control Oscillator) or DC-DC converter, to allow implementing the tuning of the Voltage or frequency, was also not implemented, and only the concept of sensor design and global and local sensor interaction was presented. Therefore, the main purpose of this work is to develop a new Controller and Sensor Tuning (CST) block, and also a digitally controlled VCO (DCO, Digital Control Oscillator), that allows on-line Dynamic Voltage or Frequency Scaling. This CST will receive information from the global sensor and from the local sensors and will decide and act over the DCO block (in fact, it can be either a DCO or a DC Voltage-Converter) to control the frequency (or voltage, if DC Voltage Converter was considered) of operation, according with the sensors' information.

The first task will be to design the CST behavior in VHDL, compile it and simulate it. Then the CST must be converted in a SPICE netlist, in order to simulate the full Power and Frequency Tuning (composed by the CST, the global sensor, the local sensors and the voltage or frequency source). The final circuit must behave with the maximum performance without any errors. Every time an error is predicted, the CST must act over the voltage/frequency source.

A second and additional objective of this work is to develop a new delay-fault correction circuit, based on the previous aging sensor circuit [19], to predict or correct the occurrence of delay-faults. The concept behind circuit failure prediction and correction has a great potential, and is a huge improve in nanometer technologies. The key idea is to propose changes upon existing aging sensors' flip-flop (FF) and to allow not only the error prediction, but also its correction, if an increase variation is achieved under additional PVTA variability.

This task starts with the design of a new reliable sensor that uses aging and PVT variability on its favor. The sensor must have low power consumption, be small, easy to be introduced locally at circuit flip-flops, where the synchronizations errors occur. A correction margin, defined by design, will allow the identification of late arrivals at FF's data input. Under certain limits, these late arrivals should be corrected, improving circuit's reliability.

SPICE simulations will be used to create sensor's sub-circuits and, also, benchmark circuits to test it under PVTA variations, in order to prove the circuit's capabilities.

1.3 ORIGINAL CONTRIBUTIONS

The original contributions developed in this research work, mainly in comparison with [25] work, are:

- Development of the CST and circuitry, in behavior VHDL representation, in structured gate level Verilog representation (using a generic CMOS library of 65nm BTM (Berkeley Technology Models)), and in SPICE type netlists;
- Specification and interconnection of a complete system, based on Frequency tuning, using a SPICE netlist description;
- Complete methodology validation, through transistor level SPICE simulation of a Circuit Under Test (CUT) and methodology circuitry;
- Development of a novel Error-Correction FF, based on previous Error Prediction FF.

1.4 CONTEXT OF THE RESEARCH WORK

This thesis research & development (R&D) work was conducted at the Superior Institute of Engineering (ISE), University of Algarve (UAlg), in close collaboration with INESC-ID Lisbon. The working group of the two Portuguese institutes has been working in partnerships with other foreigner R&D institutes and universities such as the University of Vigo is Spain, the INAOE institute in Mexico and PUCRS University in Brazil. In recent years research work has been developed in the area of aging sensors for both ASIC (Application Specific Integrated Circuit) and emulated

circuits in FPGAs (Field-Programmable Gate Array), as well as Voltage and Frequency optimization for long-term operation, both for ASIC and FPGA.

In this context, 2 *M.Sc.* thesis were already finished, and another one is currently being developed, in *ISE-Ualg.* Moreover, other M.Sc. and Ph.D. thesis were finished and are currently being developed in partner institutions, namely a Ph.D. thesis initiated in 2013 at Instituto Superior Técnico (IST) (University Partner), INESC-ID (Research Partner) and SiliconGate (Industrial Partner) in Lisbon, also in the area of Performance vs. Power optimization, but with the constraint of using commercial design tools and libraries, to allow an easier applicability to the industry.

Furthermore, this thesis pretends to continue the research work initiated in [19], to use the aging sensors developed there, along with global sensors, to optimize dynamically the Power and/or the Frequency of operation in electronic systems, and therefore develop a new DVFS (Dynamic Voltage and Frequency Scaling) methodology, compatible with PVT and aging variations.

Finally, the research work developed in this thesis was recently submitted to a definitive Portuguese Patent registration (currently pending), in August 27, 2013, under the number PT 106513, and entitled to "*Optimização ao Longo da Vida da Dissipação de Potência ou da Performance em Circuitos Integrados Digitais Síncronos*" [38]. The complete patent submission is presented in the Appendix.

1.5 THESIS OUTLINE

The remainder of this thesis is organized as follows.

In Chapter 2 the state-of-the-art in Low-Power Design Techniques are presented.

The problem of variability in CMOS nanometer technologies is addressed in the Chapter 3. The main effects, like Process, Temperature, Power-Supply Voltage, and Aging variations, that can degrade circuits' performance are resumed.

In Chapter 4 is presented how the Power and Frequency tuning is achieved. Also, it is resumed the architecture of the local sensor, of the global sensor, of the controller and of the DCO.

The Delay-Fault Correction Flip-Flop (DFC-FF) architecture is presented in Chapter 5. Based on the Adaptive Error-Prediction Flip-Flop (AEP-FF) architecture,

the DFC-FF allows the prediction of performance errors locally, in the critical memory cells where the errors occur, and the correction of the errors. The DFC-FF characteristics are also defined with SPICE simulations.

In Chapter 6, simulation results are presented, both for the Power and Frequency tuning methodology, and for the DFC-FF, using benchmark circuits as test vehicles. The results prove that the methodology can effectively tune the frequency of operation and coordinate global and local sensors to monitor Voltage variations, and that the new sensor FF has predictive fault-detection and correction capability.

Lastly, Chapter 7 summarizes the main conclusions and provides guidelines for future work.

2. LOW-POWER DESIGN TECHNIQUES

In our days, with the demand of limited energy and power consumption for mobile applications (cell phones, laptops, other battery powered devices), it became necessary to use *Low Power* (LP) design techniques for successful product development. *V_{DD} reduction* and *clock frequency reduction* are two techniques to reduce power dissipation in CMOS. However, certain product's specifications may not comply with an overall performance loss.

Nowadays there are several low power techniques. Among dynamic power control techniques [1] are clock gating, multi power supply voltage (multi-V_{DD}), variable frequency, and power efficient circuits. Leakage power control techniques comprise power gating and multi threshold (multi-V_{th}) cells. Depending on the design complexity and requirements it could be adopted a combination of LP techniques.

This section presents a brief description of few LP design techniques used to reduce power consumption in modern *Systems on a Chip* (SoC) designs. For the concern of this thesis, we will focus on DVS and/or DVFS techniques.

2.1 DYNAMIC VOLTAGE AND FREQUENCY SCALING (DVFS) FUNDAMENTALS

Dynamic voltage scaling (DVS) is a widely used technique to reduce the overall energy consumption of a processor, especially under wide workload variations [8]. *Dynamic* multi-V_{DD} design strategies can be broadly classified as follows [1]:

- *Static Voltage Scaling (SVS):* Different, but static supply voltages are applied to different blocks or subsystems of the SoC.
- *Multi-level Voltage Scaling (MVS):* The block or subsystem of the ASIC or SoC design is switched between two or more supply voltage levels. However,

for different operating modes, a limited number of discrete voltage levels is supported.

- *Dynamic Voltage and Frequency Scaling (DVFS):* Supply voltages, as well as clock frequency, are dynamically varied for different operation modes, in order to improve power efficiency. When high speed of operation is required, V_{DD} is increased to attain higher speed of operation; of course, the penalty is increased power consumption.
- *Adaptive Voltage Scaling (AVS):* In this design strategy, the supply voltage is controlled using a control loop. This is an extension of DVFS.

The average dynamic power dissipated by a chip using static CMOS gates is $P_{ave} = P_{dyn} = \sum_{ni}(\alpha_i \cdot C_i \cdot f_{clk} \cdot V_{DD}^2) = C \cdot V_{DD}^2 \cdot f_{clk}$, where f_{clk} is the clock frequency, C_i is the output node capacitance ni of each logic element, and α_i is the average switching activity parameter of node ni. C is the capacitance being switched per clock cycle. [1]

Due to the quadratic dependence of energy with supply voltage, significant energy savings are achievable with DVS. However, this formula is only accurate with chips implemented using 100% static CMOS. Also, some electrical components' efficiency decrease with increasing temperature. Hence, power consumption will increase to compensate the temperature rise. Consequently an increase in V_{DD} or f_{clk} may increase system power demands even faster than the CMOS formula indicates, and vice-versa [1].

In devices powered by batteries, DVS is widely used as part of strategies to manage switching power consumption. This leads to low voltage modes in conjunction with lowered clock frequencies to minimize power consumption used in components such as CPUs (Central Processing Unit) and DSPs (Digital Signal Processor). When computational power is needed the voltage and frequency will be raised.

There are other power conservation techniques that use the same principles as DVS, and that technique is *Dynamic Frequency Scaling* (DFS). Both techniques can prevent overheating in the computer system, however if the voltage supplied is reduced below the limit recommended by the manufacturer, it can result in system instability.

Even though DVS, DFS or DVFS may reduce significantly power consumption, it is important to tune V_{DD} and f_{clk}, otherwise the desired performance will be sacrificed. Nevertheless, the major difficulty is to find the "correct" values for V_{DD} and f_{clk}. This occurs because is difficult (1) to know the relationship between V_{DD} and f_{clk}; and (2) to have a circuit that matches the critical path [1].

Enhanced Intel SpeedStep® Technology [9], AMD PowerNow!™ technology [10] and *Razor* [8] [12] from ARM are examples of DVS/DVFS techniques. Those techniques choose the "correct" V_{DD} / f_{clk} value based on the statistic of errors obtained. The next sub-sections will consider those techniques.

2.2 RAZOR

2.2.1 RAZOR I

Accordingly with [8], Razor (so on called Razor I) key concept is to sample the input data of the flip-flop at two different points in time. The first sample is stored in a conventional positive-edge triggered, master-slave flip-flop. This main flip-flop has a "shadow latch", which samples at the negative edge of the clock. Thus, the shadow-latch gets additional time equal to the high-phase of the clock to capture the correct state of the data. When the data at the main flip-flop differs from the shadow-latch data an error is flagged.

Because it is allowed to violate the setup and hold constraints, it is required an additional detector to flag the occurrences of metastability at the output of the main flip-flop.

Error signals of individual Razor I flip-flops are then "OR"-ed together to generate a pipeline restore signal which overwrites the correct data in the shadow-latch into the main flip-flop, thereby restoring correct state in the cycle following the errant cycle. [8]

Because the shadow latch data will be used to overwrite state in the main flip-flop, it is required to ensure that data is always correct, and for that it will be necessary to use conventional worst case techniques.

However, deploying Razor I for high-performance, aggressively-clocked microprocessors, creates a few key design issues. The main issue is the generation and propagation of the pipeline *restore* signal [11][12]. This signal is evaluated at the output of a high fan-in OR-tree and is suitably buffered and routed to every flip-flop in the pipeline stage before the next rising edge of the clock. But, this can cause the error recovery path to become itself critical when the supply voltage is scaled due to timing constraints on the *restore* signal. Hence, Razor I is used especially in aggressively clocked designs, as limited voltage headroom is available. Other issue is the design of the metastability detector. Due to rising process variations, it becomes difficult to respond to metastable flip-flop outputs across all process, voltage and temperature corners.

Consequently, it is necessary the use of larger devices that will impact adversely in the area and power overhead of the Razor I flip-flop. Also other risk is the metastability at the *restore* signal which can propagate to the pipeline control logic, leading to a potential system failure.

In order to address the design and time issues of Razor I, it was developed an improved technique called Razor II. This technique will be considered in the next sub-section.

2.2.2 RAZOR II

Razor II technique [12] moves the responsibility of recovery entirely to the micro-architectural domain. In Razor II, the processor is intended to operate near the Point of First Failure (PoFF) and recovery from a timing error occurs by a conventional architectural replay mechanism.

Razor II is significantly more amenable to high performance microprocessors due to architectural replay that greatly simplifies the error recovery path, when compared with Razor I.

The Razor II approach is based in two components which are as follows [12]:

1. Razor II only performs detection in the flip-flop, leaving correction to be performed through architectural replay. This allows reduction in the

complexity and size of the Razor II flip-flop, even though at the cost of increased Instructions Per Cycle (IPC) penalty during recovery. Architectural replay is a conventional technique which supports speculative operation such as out-of-order execution and branch prediction. Consequently, in the event of timing errors, it is possible to overload the existing framework to support replay. Additionally, this technique excludes the need for a pipeline *restore* signal.

2. The design of the Razor II flip-flop replaces the master-slave flip-flop by a positive level-sensitive latch. The flip-flop will flag any transition on the input data in the positive clock-phase as a timing error. Without the master latch, the clock-pin capacitance of the flip-flop is reduced and reducing also the power and area overhead. This allows Razor II flip-flop, without additional overhead, to detect Single Event Upsets (SEU) in the logic and registers.

In the Razor II flip-flop, the elimination of the master latch and the metastability detector are major simplifications that lead to improvements in delay, power and area. The elimination of the master latch leads to slightly improved clock delay compared to a conventional flip-flop and eliminates completely the setup time constraint at the positive edge of the clock. The total power overhead due to insertion of Razor II flip-flops in the processor is around half the power used with Razor I. The area improvement is due to completely elimination of the metastability risks at the positive edge of the clock for the latch data-path, thereby eliminating the need of a metastability detector.

2.3 INDUSTRY STANDARDS

2.3.1 ENHANCED INTEL SPEEDSTEP® TECHNOLOGY

Enhanced Intel SpeedStep® Technology is mean to achieve average power savings depending on system usage and design.

This technology [9] allows the system to adjust dynamically the processor voltage and core frequency, which can result in decreased average power consumption and

decreased average heat production. Furthermore, this feature can help companies with power concerns due to their sites approaching the limits of bounded electrical infrastructures. Enhanced Intel SpeedStep® Technology, combined with existing power saving features, can provide an excellent balance between conserving power when you do not need it and providing it when you do.

Enhanced Intel SpeedStep® Technology architecture uses design strategies that include [9]:

- Separation between Voltage and Frequency changes. By stepping voltage up and down in small increments separately from frequency changes, the processor is able to reduce periods of system unavailability (which occur during frequency change). Thus, the system is able to transition between voltage and frequency states more often, providing improved power/performance balance.
- Clock Partitioning and Recovery. The bus clock continues running during state transition, even when the core clock and Phase-Locked Loop are stopped, which allows logic to remain active. The core clock is also able to restart far more quickly under Enhanced Intel SpeedStep® Technology than under previous architectures.

Because the latency associated with changing the voltage/frequency pair (referred to as P-state) is reduced by Enhanced Intel SpeedStep® Technology, those transitions can be practically undertaken more often, which enables more-granular demand-based switching and the optimization of the power/performance balance based on demand [9].

2.3.2 AMD POWERNOW!™ TECHNOLOGY

AMD PowerNow!™ technology [10] is an advanced power-management feature that reduces the overall power consumed by the processor through control of voltage and frequency. This power-saving is achieved by controlling independently the voltage and the frequency. This technology is designed to provide very fine and dynamic control of voltage and frequency.

AMD PowerNow!™ technology [10] have tree modes of operation: High-Performance (where the processor runs at the maximum rated voltage and speed providing the maximum performance possible);. Power-Saver (where the processor runs at the lowest voltage and frequency supported); and Automatic (where the speed and voltage are dynamically and automatically determined by actual performance demands of the application). In the Automatic mode, is the application environment needs that dictates the amount of performance required and power used. This operation mode guarantees that only the essential amount of power is dissipated to meet the performance demands of the application.

3. VARIABILITY IN CMOS CIRCUITS

Nowadays the demand for functional circuits is increasing. Also the reduction of size in the SoC demands that the causes of performance errors to be well know.

In this section, several causes of performance errors are addressed. A detailed analysis of all the causes of performance error is out of the scope of this thesis. The idea is to give an overall view in order to the reader understand the importance of this research work. With this in mind, and in case the reader wants exhaustive information, he is invited to obtain detailed information by consulting the references.

3.1 AGING

Aging is the process where circuits lose their initial characteristics. It is in general a change in the behavior of all the transistors of an integrated circuit. This modification which is slowly progressive and different from transistor to transistor, usually translates into a decreased ability of conducting electric current (in the case of BTI corresponds to the increase in the absolute value Vth of the transistors). However, aging is not occasional random deterioration where a component is damaged (often caused by extreme variations in the manufacturing process) or deterioration caused by operating in special situations of stress, causing irreparably damage to a component (for example, a condenser disrupting or electrical connections breaking in a component) [3].

Also, the constant size reduction in CMOS technology boosts the aging of integrated circuits. Smaller transistor's, where the driving is done by a small number of ions and electrons, turns the circuits and components more vulnerable to external influences. Variation of any existing parameter in nanotechnologies leads to greater uncertainty in the behavior of electrical circuits and to lower energy needed to modify

this behavior. This way the circuits became more susceptible to changes in its structure and in its behavior. Aging is thus accelerated in nanotechnologies [3].

In earlier technologies, this aging also occurred, but as the effects on the circuits were smaller, they weren't notice during the life cycle of the circuit. The life cycle of the circuit were guaranteed by generous safety margins (time slacks). At that time it was said that the circuits practically didn't age.

However, in nanotechnology, and more specifically in the case of digital integrated circuits, aging usually causes the increase in propagation time of the logical gates and, thus, the propagation times of the paths. If the propagation time increases, a synchronous error may occur when the time slack is not able to accommodate this increase in propagation times inside the period of the clock. In this case, the flip-flops (and the synchronous memory cells) do not capture the correct value because it arrives late in relation to the clock signal (which marks the synchronism) [13]. The existence of PVT (Process, power-supply Voltage and Temperature) variations in conjunction with the cumulative degradation caused by aging (ie, overall PVTA variations) does reduce the useful lifetime of circuits. In fact, as reported in [14], considering nanotechnology applications where safety is critical, and where the circuits average time of use are relatively high, as is the case of automotive electronics where trucks have average life time of 15 to 20 years (and where nanotechnologies is the technology frequently used [15]), the cumulative degradation caused by aging is critical for proper circuit performance. It is imperative that aging is considered in the design stage, in order to predict and ensure circuits operationally during the lifetime expected without errors, especially in new nanotechnologies [16].

The most important aging causes that cause degradation in the digital IC are [3]: Bias Temperature Instability (BTI); Hot-Carrier Injection degradation (HCI); Time Dependent Dielectric Breakdown (TDDB); and Electromigration (EM).

From all these phenomena, the BTI stands out as being the one that causes more degradation in ICs affecting the transistors.

3.1.1 BIAS THERMAL INSTABILITY

In technologies under 130nm, Negative Bias Temperature Instability (NBTI) in PMOS devices has become an important reliability issue [17]. NBTI consists in the increase of the absolute threshold voltage value with time in the PMOS transistors. This will cause the circuit delay to degrade over time until the delay exceeds the circuit specifications [17].

There is also a corresponding effect in the NMOS transistors, identified as Positive Bias Temperature Instability (PBTI), and is observed when a positive bias stress is applied across the gate oxide of the NMOS device [17]. Even though PBTI impact is lower than NBTI, with the use of Hf-based dielectrics in the gate-oxide for leakage reduction, PBTI is gradually becoming more important [17].

BTI is a thermally activated process that degrades CMOS transistors' physical parameters like threshold voltage and channel holes/electrons mobility. It occurs under negative gate voltage for PMOS (e.g. VGS = -VDD) and under positive gate voltage for NMOS (e.g. VGS = VDD) and is measured as a gradual increase in the magnitude of threshold voltage (|Vth|) with time. The result of the rise in |Vth| is the reduction of the absolute drain current (I_{Dsat}), the reduction of the trans-conductance (g_m), and the increase of absolute drain off current (I_{off}). The drain current reduction outcomes in temporal degradation in the performance of a circuit causing reliability degradation over time [19].

In a digital circuit, a PMOS transistor is conducting normally when the gate is with the logic level "0", which is equivalent to have $V_{GS} = -V_{DD}$ (an NMOS transistor is conducting normally when the gate is with the logic level "1", which is equivalent to have $V_{GS} = V_{DD}$). In this condition, also known as the stress stage of the transistor, the electric field employed between the gate and channel transistor has a maximum intensity. Furthermore, if a greater thermal energy (high temperature) is applied, the conditions to break the chemical bonds (Si bonds and Si-H in the interface SiO2) are boosted, increasing the instability and creating obstacles to the drive (which reduces the mobility of holes for the PMOS or electrons for the NMOS, causing a reduction of the absolute drain current, I_D). In the condition where the logic level of the gate is high for the PMOS ($V_{GS} = 0V$ and $V_G = V_{DD}$) or low for an NMOS ($V_{GS} = 0V$ and $V_G = 0V$), some of the Si-H bonds are restored, so this is also called the recovery

phase. However, not all connections are re-established, which leads to a slow and cumulative degradation of the transistor (Figure 3.1). That is, considering the transistor as a switch (as in digital circuits), the time between the cutting and driving, and vice versa, will be higher [3].

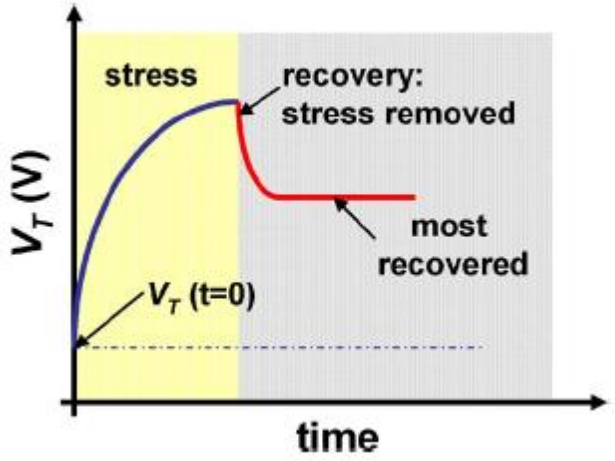


Figure 3.1 - PMOS NBTI versus time illustrating both degradation and recovery.

Therefore, the BTI aging mechanism is causing most damage in digital IC and is currently the greatest concern to the scientific community in the field of nanotechnology and therefore the most investigated. The increase in the transistors |Vth| causes reduction of the useful life of the CMOS. This happens due to the increase of digital logic gates delay times and paths response times which leads circuits to begin to fail and, consequently, all the equipment in which they are inserted [3].

It is important to note that the NBTI effect affects the PMOS devices, and hence the rising delay of a gate, while the PBTI effect affects the NMOS transistors, and therefore the falling delay only [17]

There are several approaches to guard-banding a circuit and guaranteeing optimal performance over its lifetime, such as sizing, and synthesis, which can be categorized as “one-time” solutions that enhance appropriate guard-bands at design time [17]. This type of approaches results in huge positive slack times during the first years of operation of the circuit, and so a greater than needed power and area overhead, in comparison with a circuit designed to meet the precise specifications, at every time point, during its operation lifetime [17]. Also, adaptive control systems can be

implemented using a critical path based approach. Those techniques may be used to counter the impact of aging, as well as process and temperature variations[17].

3.1.2 OTHER AGING EFECTS

Although the BTI is the dominant effect on the aging process of the CMOS transistors and integrated circuits, other factors also influence the performance degradation of the ICs. The most important are discussed in the following.

3.1.2.1 HOT CARRIER INJECTION

Besides BTI, Hot Carrier Injection (HCI) is one of the aging effects that most degrades IC's normal operation. Similarly to BTI effect, HCI affects the transistor transitions speed between the conduction and cut states. HCI might be a limitation on transistors size reduction, since this effect is strongly dependent on design, material quality and processing parameters of the MOSFET [3]. Despite the degradation impact of HCI is less critical in comparison with NBTI, the transistors size reduction (considering the length of FETs gate), turns HCI into a mechanisms that must be considered.

The HCl phenomenon is the degradation of the threshold voltage (Vth) and mobility of the FETs under dynamic stress mode. The electrons presents in the channel that moves from source to drain are subjected to an electric field with variable intensity depending on the bias conditions, resulting in some of the same electrons being injected into the gate oxide of the transistor [3][19]. Thus, the dielectric property of the oxide ends up being affected, and this causes a reduction in the NMOS and an increase in the PMOS current capacity [18]. Consequently, the operating speed of the circuit is reduced, what could potentially cause failure during long term operation.

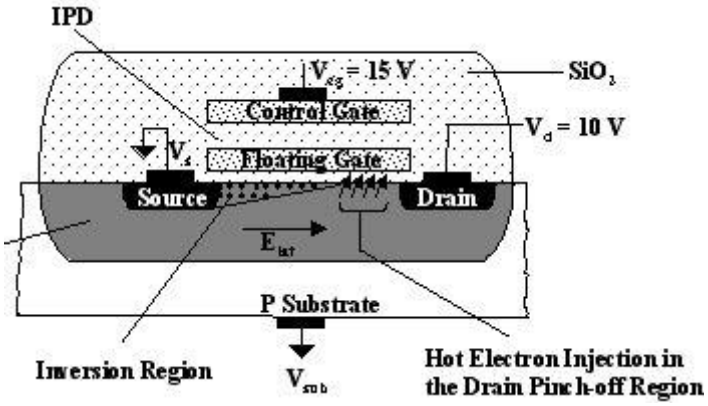


Figure 3.2 - Hot carrier effects.

This type of degradation becomes more important due to transistors' size reduction, especially on the length of the FETs channel. As the channel length of the FET is reduced, higher switching speed is obtain, but this leads to an increase in the electric field present in the channel. This increase in the electrical field causes physical changes on the FETs gate oxide due to deposited charges therein resulting in device performance degradation [3].

HCI have a higher impact on NMOS transistors, and this form of degradation is more pronounced near the drain region and is enhanced at saturation [19][5].

3.1.2.2 TIME DEPENDENT DIELECTRIC BREAKDOWN

The Time Dependent Dielectric Breakdown (TDDB) degradation is a failure in the transistor gate oxide. This failure happens when a conductive path forms between the gate and the substrate through the oxide, and turn out to be impossible the control of the current flow between drain and source [1][19].

TDDB can be classified as hard breakdown (HBD) or soft breakdown (SBD). HBD is a catastrophic failure of the device and of the entire circuit, while SBD does destroy the transistor functionality over time. In the SBD the dielectric properties are maintained, but there is an increase of the gate current. SDB is responsible by parametric variations such as energy, delay and noise margin over time [1][19].

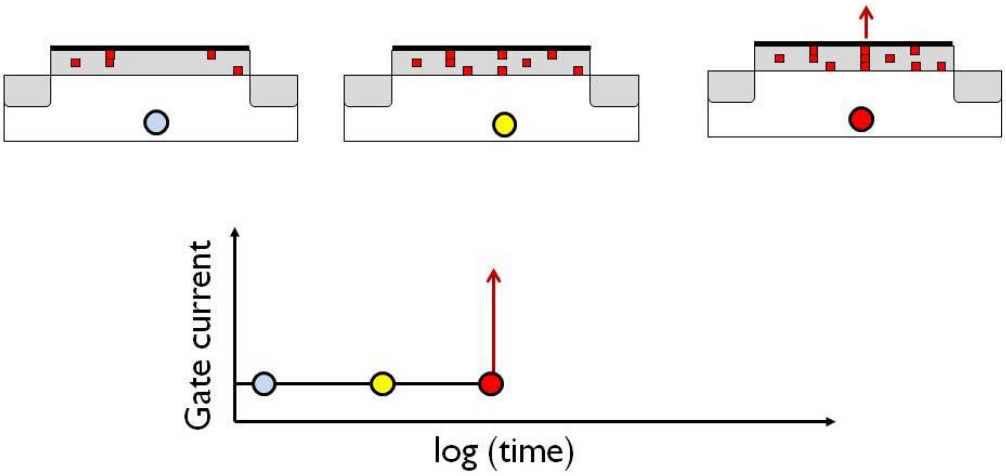


Figure 3.3 - TDDB effect cross section view.

According to [19], the gate oxide break down is sensitive to junction temperature, gate voltage and gate oxide area. TDDB affects in a higher grade MOSFET devices when they are ON. As a result of scaling, MOSFET devices with oxide thicknesses below 2nm, became more predisposed to TDDB effects.

3.1.2.3 ELECTROMIGRATION

Electromigration (EM) is the gradual displacement of metal atoms in the semiconductor conductive interconnections. It is a reliability problem in CMOS devices, because this effect can cause breakage of conductors, leading to circuit failure due to open circuits or short-circuits.[3][19].

EM is a function of current density and it is also accelerated by elevated temperature. This effect happens due to the direct action of the electric field on the charged atoms or ions of the metal and the frictional force or momentum exchange between the flowing electrons and these ions [20].

Metal films have imperfections or microstructural variations that cause non-uniformly distributed atomic flow rates through them. This flow rates, or flux divergence through different sections of the conductor, result in mass depletion causing voids and mass accumulation causing hillocks [20].

EM is influenced by temperature, current density, wire length and material. It is also boosted by higher temperature and current density [19]. However, the acceleration effect of high temperature on EM becomes emphasized only when a void has started to form in the metal line. Before any void is formed, the metal can still be under uniform thermal distribution [20].

Also, the resistance of a wire exposed to EM will increase with time, until a void appears. As a consequence, the RC time constant on the data line will raise, leading to gradual propagation time degradation. With nanometer technologies, high density currents are predictable, because dimension scaling was performed without proportional current reduction, what makes the transistors very vulnerable to EM induced failure [19].

3.2 POWER-SUPPLY AND TEMPERATURE

Not only aging (A) affects circuit's performance, but also process variation (P), power-supply voltage (V) and temperature (T) variability, the so called PVTA variability. It is interesting to note that despite these three effects contribute to circuit's aging they also affect the performance of the circuits by itself [19].

When systems are designed, they are design to operate at a nominal supply voltage. However, for many reasons, this voltage may vary. Furthermore, the supply voltage on a circuit is not distributed uniformly and varies in time and across the chip [19]. Among the reasons for supply voltage variations are the tolerance of the voltage regulator, IR drops along supply rails due to the connection lines resistance, and di/dt noise [22]. While a system is being designed, the engineer/designer team can trade-off power supply noise against resources dedicated to power supply regulation and distribution; usually the supply is defined with a range of more or less 10% around nominal value at each logic gate [22]. Because speed is proportional to VDD, this will lead to a delay variation in approximately the same order.

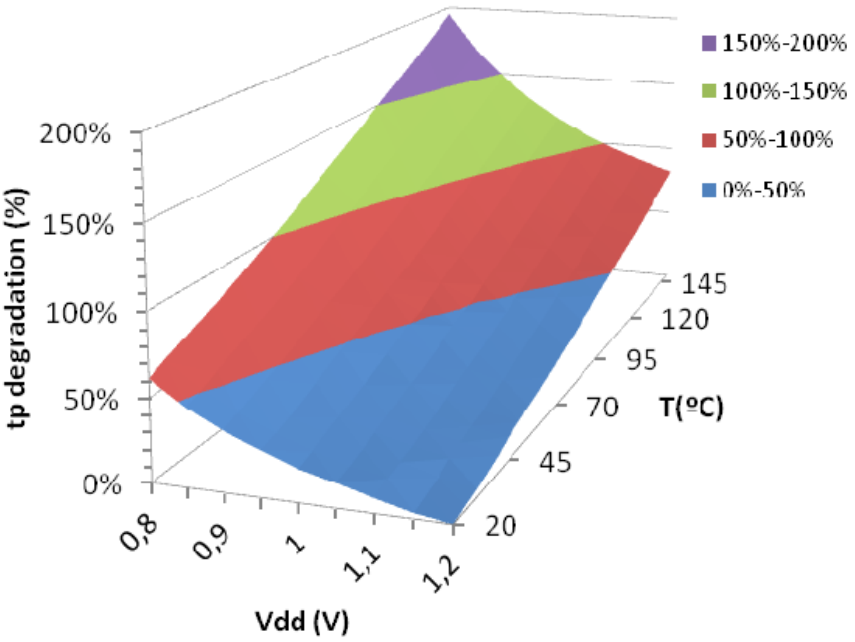


Figure 3.4 - 65nm performance vs temperature and supply voltage variability. Courtesy of [19].

Regarding power-supply voltage variations' impact in circuit's propagation time, C. Martins [19] made a study where it was used a 20x inverter chain test circuit, with 65nm Berkley PTM technology and simulated in SPICE. The propagation time degradation was up to near 70%, when t power-supply voltage decreased from 1.2V to 0.8V and considering a temperature of 27ºC (Figure 3.4).

The junction temperature of a transistor is the temperature at the semiconductor junctions that forms the transistors and is defined as the sum of the ambient temperature and the temperature rise caused by power dissipation. The rise of this temperature is determined by the power consumption and the thermal resistance. As consequence, when the temperature increases, the drain current decreases [22].

The integrated circuits must function under a range of ambient temperature according with its use. Common temperatures ranges, depending on use, are defined in Table 3.1.

USE	Minimum Temperature (ºC)	Maximum Temperature (ºC)
Commercial	0	70
Industrial	-40	85
Military	-55	125

Table 3.1 - Ambient temperature ranges standards.

Despite the maximum ambient temperature for commercial use is 70ºC, is common to see circuits operating with junction temperatures up to 125 °C.

Due to different operation activity and electrical physical characteristics on the circuit blocks of a chip, the power dissipated and the junction temperature are not uniform across the chip's die (Figure 3.5), having as consequence different performance degradation [19].

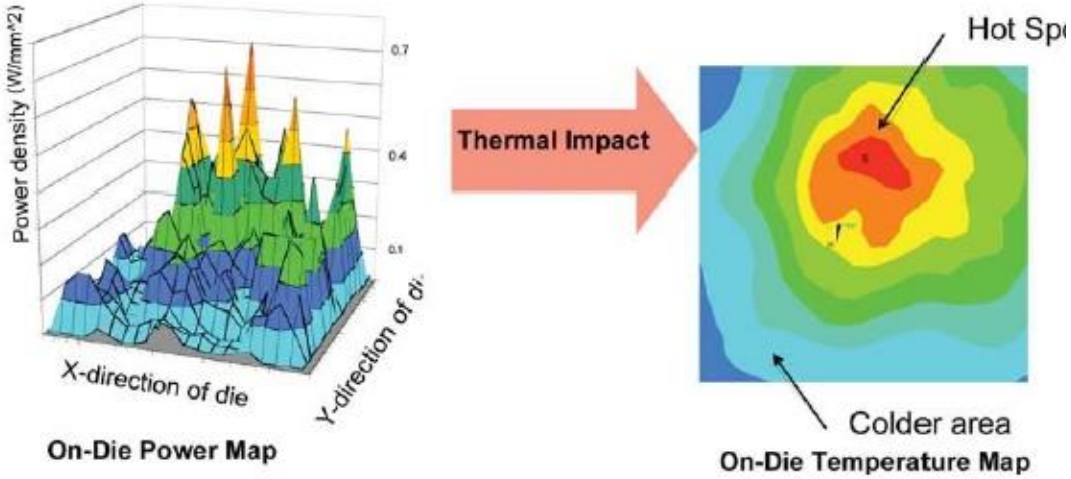


Figure 3.5 - Junction temperature across the chip's die.

Also, temperature variations' impact in circuit's propagation time was study in [19]. Figure 3.4 shows results for the 20x inverter chain test circuit, with 65nm Berkley PTM technology, used and simulated in SPICE. The propagation time degradation was up to near 100%, when temperature increased from 20ºC to 145ºC and considering a supply voltage of 1.1V (Figure 3.4). Considering both effects accumulated (temperature and power-supply voltage) the propagation time degradation was near 200%.

3.3 PROCESS VARIATION

While technologies continue to pursuit Moore's Law, a variety of challenges will rise. One of these challenges that need to be overcome is management of process variation [32]. Process variation has always been a critical aspect of semiconductor fabrication.

In 1961, Shockley, was the first to discuss random variation in semiconductor devices by analyzing random fluctuations in junction breakdown. In 1975, Keyes extended to MOS devices Shockley's concepts of random variation, by modeling the effect of random fluctuations in the number of impurity atoms in the depletion layer of a field-effect transistor (FET) [32].

Due to systematic variation in MOS devices, Schemmert and Zimmer, in 1974 computed the sensitivity of ion-implanted MOS threshold voltages as a function of the implantation energy and the oxide thickness.

In 1980, Yokoyama presented an extensive analysis of threshold voltage sensitivity using a closed-forum numerical simulation with a Monte Carlo approach. In the next years, several authors expand their research.

Process variation can be defined as circuit's geometry and parameter variations in devices and interconnects due to manufacturing variability. Small variations in the transistors effective channel length or/and in the gate oxide film thickness may outcome in variations in the threshold voltage and leakage current, having an effect similar to aging reducing the circuit's performance. The difference between process variation performance degradation and aging performance degradation is that process variation performance degradation is static, while aging performance degradation changes over time [19].

Also, variations in interconnects (e.g. contact resistance variations plus metal layers width, thickness, spacing variations) are responsible for coupling capacitance variations and change global line resistance, causing in lines voltage drop, nodes capacitance and crosstalk variability [19].

In the same die, depending on the position on the wafer, performance variation is observed due to process variability. Performance variation may be as significant as

30% [19] due to process variability. That's why the continued effort for decreasing the size of semiconductors means that management of variation will play a significant role in future technology scaling.

In the next subsections the main critical sources of variation will be addressed.

3.3.1 CRITICAL SOURCES OF VARIATION

Some of the process variation effects are [32]:

⇒ Random Dopant Fluctuation (RDF);

⇒ Line-Edge Roughness (LER) and Line-Width Roughness (LWR);

⇒ Variations in the Gate Dielectric:

- Oxide Thickness;
- Fixed Charge;
- Defects and Traps;
- Patterning Proximity Effects:
  - Classical;
  - Associated with OPC;
- Polish:
  - Shallow Trench Isolation (STI);
  - Gate;
  - Interconnect;
- Strain:
  - Wafer-Level Biaxial;
  - High-Stress Capping Layers;
  - Embedded Silicon-Germanium (SiGe);

- Implants and Anneals:
  - Tool-Based;
  - Pocket Implants;
  - Rapid-Thermal Anneal (RTA);
  - Poly Grains.

3.3.2 RANDOM DOPANT FLUCTUATION (RDF)

As the number of dopant atoms in the channel decreases with scaled dimensions, the impact of the variation associated with the atoms increases [32]. Also, MOS threshold voltage variation due to RDF increases, as the device area decreases.

The major contributor to device mismatch of identical adjacent devices is assumed to be RDF. In practice, it is known that additional effects also contribute to the measured variation and identifying the magnitude and root cause for these additional effects is important in facilitating the development of mitigation techniques [32].

3.3.3 LINE-EDGE ROUGHNESS (LER) AND LINE-WIDTH ROUGHNESS (LWR)

The primary concern with LER/LWR is variations in poly-gate patterning. LER and LWR are associated with increases in the sub-threshold current as well as degradation in the threshold voltage (Vth) characteristics [32]

Accordingly to Fukutome [32], the roughness of extension edges induced by gate LER depended on the implanted dose, halos (pockets), and various co-implantations. With a decrease in the average LER, there is an improvement of 4nm in Vth roll-off, and co-implants induced a degradation of 5mV in the standard deviation of Vth.

Asenov [32] studied the combined effect of LER and random discrete dopants on current fluctuations and concluded both acts in a statistically independent manner.

Also, LER-induced current fluctuations have much stronger channel length dependence and, as devices are scaled to shorter dimensions, LER is expected to supplant RDFs as the dominant variation source.

3.3.4 VARIATIONS IN THE GATE DIELECTRIC

The gate dielectric variations include variations in oxide thickness, fixed charge, and interface traps. These physical changes in the dielectric result in parametric variations in drive current, gate tunneling current, or threshold voltage [32].

3.3.4.1 OXIDE THICKNESS

Asenov [32] studies show that intrinsic threshold voltage fluctuations induced by local oxide thickness variations can be comparable to voltage fluctuations equivalent to variations caused by RDF for conventional MOS devices with dimensions 30nm and below.

3.3.4.2 FIXED CHARGE

The presence of fixed charge can affect the mobility and the threshold voltage. As a consequence, variation in the fixed charge may affect the uniformity of the threshold voltages on devices.

3.3.4.3 DEFECTS AND TRAPS

Fast transient charging (FTC) in electron traps can cause electron mobility degradation and Vth instability.

Also two types of defects have been detected, those associated with grain boundaries in the nano-crystalline HfO2, and those associated with different charge states of the O-atom vacancy [32].

3.3.4.4 PATTERNING PROXIMITY EFFECTS

Lithographic measure of aggressiveness includes illumination conditions, resist materials/chemistry, optical proximity correction (OPC) and other resolution enhancement techniques (RETs).

A variety of techniques can be applied to layers to improve the lithographic patterning and reduce the variation. One of the most powerful of these is OPC. [32]

Accordingly with [32], OPC pre-distorts the mask data following specific algorithms in order to achieve a desired pattern on the wafer. OPC is based on process model that incorporates lumped optics, resist, wafer stack, and mask effects. This model generates a mask-to-wafer optical transfer function, and an OPC algorithm is written to invert the transfer function. An OPC formula is developed using an iterative algorithm that modifies the starting database in order to achieve the desired pattern on the mask.

3.3.4.5 POLISH

A critical process step in advanced semiconductor technologies is chemical mechanical polish (CMP). This step is very important; hence, CMP can create a defective semiconductor. That is why recent literature devotes attention to modeling interconnect variation.

For instance [32], Yu et al. characterize the smoothing and planarization effects of ILD (Inter-Layer-Dielectric) polishing, Soumyanath [40] present a technique based on simple time-delay measurements from a repetitive waveform and Mehrotra [41] analyze interconnect timing performance in a high-speed microprocessor by using timing analysis in conjunction with a post-extraction net adjustment.

3.3.4.6 STRAIN

Researchers explore channel strain approaches for transistor enhancement. In the beginning of the 21st century, a new class of transistor strain approaches was developed, that used process features external to the transistor (rather than strain in the channel itself as with the biaxial approaches) to strain the transistor [32].

New variation challenges arise, both random and systematic due to process strain. For instance, Tsang [32] developed an analytical model to predict threshold variation as a function of Ge fraction, layer thickness, channel length, and doping profile. Simulations and experimental data for n-MOSFETs and p-MOSFETs in both single and dual-channel architectures verified this model.

3.3.5 IMPLANT AND ANNEAL

There are a number of variation sources associated with the physical implant and anneal processes. The implant tool conditions are a significant source of transistor variation.

Devices variations can be affected by [32] accuracy of dose, purity of dose, spike anneals peak temperature, the ramp-up and cool-down rates, the architecture of the pocket (halo) and extension (tip) implants. Also the advanced Rapid thermal Annealing (RTA) processes have introduced new variation sources.

Another variation mechanism related to implant technology ascends from the poly-crystalline nature of conventional gates. Higher diffusion, implant channeling along grain boundaries and variations in dopant activation, can all cause increased variation.

Fukutome [32] study show that by optimizing the grain boundary they were able to get a 26% reduction in threshold voltage variation and Brown (referred in [32]) concluded that if LER does not scale by the International Technology Roadmap for Semiconductors (ITRS), the dominant source of variability for channel lengths below 25nm it will be fluctuation due to poly-grain boundaries.

3.4 EFFECTS OF VARIABILITY IN PERFORMANCE DEGRADATION

Synchronous digital circuits operation is controlled by a clock signal. For these circuits, maximum operating frequency is the standard metric to quantify their performance. The slowest combinatorial logic path between two memory elements, also known as critical path (CP) limits a circuit's clock frequency on a digital circuit [19].

Synchronous digital circuits are sequential circuits usually built with several flip-flops or latches, interconnected with combinational logic and sharing a common clock signal. To assure proper operation, flip-flops require that signals be stable at their inputs a minimum time before and after the clock edge, known as setup times (t_{setup}) and hold times (t_{hold}), respectively. High performance operation will then require that the circuit works with the maximum clock frequency possible. In this condition, any extra delay in the critical path will transfer signal transitions into the setup time window, making the flip-flop to capture an unpredictable value and eventually cause circuit's operation malfunction and failure, referred as a Delay Fault [19].

A circuit path subjected to variability will suffer from additional delay (i)[19]. Aging mechanisms will degrade circuits, turning combinatorial paths slower with time. When temperature and power-supply voltage operation conditions drift away from normal to worst case conditions (higher temperature and lower voltage), paths' delays will also increase. Also, manufacturing process will introduce uncertainty on the nominal circuit delays between equal chips, from the same wafer or from different wafers. Usually, the additional delays introduced by variability are analyzed at design time, and are accounted as an extra time (time slack) when selecting the minimum clock period, to keep circuits reliable. However, this additional time slack adds performance penalties on the circuit, reducing the maximum clock frequency [19].

During circuit operation it is expected that each path will suffer delay time variations induced by PVTA variations. These variations are not proportional on every path, and the initial delay ranking may be modified, with a new path becoming the critical path [19].

4. POWER AND FREQUENCY TUNING

The purpose of this chapter is to define and present the complete aging-aware power or frequency optimization methodology. The methodology is based on both Global Sensors and Local Sensors, which are used to allow a constant and on-line performance monitoring, and to allow an optimized operation, targeting reduced power consumption or high performance. Performance optimization is achieved by increasing operating frequency to the maximum limit while still preventing errors from happening. Global sensors are broad-spectrum and must be tuned according to local sensor failure prediction (off-line). Local sensors are placed in judiciously selected memory cells, to monitor critical paths (CPs) and nearly-CPs. Moreover, local sensor design and insertion is performed using an aging prediction analysis, to determine the infant and the aged delays of the CPs (which may differ). Next subsections will explain the methodology in detail.

4.1 GLOBAL SENSOR

The objective of the global sensor is to perform a periodic on-line monitoring of PVTA variations over the Circuit Under Test (CUT). Because is not possible to replicate each Critical Path (CP) time variation of the CUT, it is used a dummy CP instead. This dummy CP is a “pessimistic” signal path and not a replica of a CP of the CUT. So, the dummy CP ages faster than the real signal paths of the CUT.

Additionally, the dummy has different outputs along the path, creating different propagation delays at each output to match the probable CUT’s CP delay, in accordance with predictable PVTA variations [25]. Also, the global sensor was designed in order that the activation period of the sensor affects the aging degradation of the dummy. So the less time the sensor is activated, greater is the aging

degradation, also with the benefit of saving power and provide a safety margin to CUT's CPs time slacks.

Also with the CUT off-line, it is possible to tune the global sensor with local sensor activation, reducing the difference in critical paths' propagation delays (the dummy's and the CUT's CP) [25]

4.1.1 ARCHITECTURE

Figure 4.1 presents the global aging-aware performance sensor (GAAPS) architecture, with two CP delay's replicas implemented with NOR gates and with NAND gates.

With the knowledge of the critical paths of the circuit, it is possible to create two fictitious paths, with propagation times higher than the critical paths expect during the lifetime of the circuit. One chain is implemented with NOR gates creating a fictitious critical path which will, presumably, age more than the critical paths of the circuit when subjected to NBTI effect (which strongly influences the degradation of Vth of PMOS transistors). The other chain is implemented with NAND gates, creating another fictitious critical path which will, presumably, age more than the critical paths of the circuit when subjected to the PBTI effect (which strongly influences the degradation of Vth of NMOS transistors).

There is a control unit to process the data received from the 2 CP dummies. The control unit receives a signal to start the analysis of the circuit's performance and generates control signals ($Test\_data\_1$, Age_Enable_1, $Test\_data\_2$, $\overline{Age\_Enable\_2}$) to operate overall performance analysis. These control signals allows to place transparent NOR and NAND gates chain inputs, through the signals (Age_Enable_1, $\overline{Age\_Enable\_2}$), so that through the signals ($Test\_data\_1$, $Test\_data\_2$) a test sequence is generated that stimulates the two state transitions at the outputs of the gates of the two chains.

Along the two chains some cells will be connected to the output of several NAND and NOR gates signals. These cells create several fictitious paths with different propagation times. The architecture of the global sensor cells is defined in [38] and in [25], and is addressed in section 4.1.3. The elements (D-type Latch, activated in the

low state of the clock signal, a Delay Element and a Stability Checker) connected to the output of the NAND and NOR gates signals are already defined in [37]. The outputs of the NAND and NOR L+DE+SC blocks sensors form two outputs sensor data.

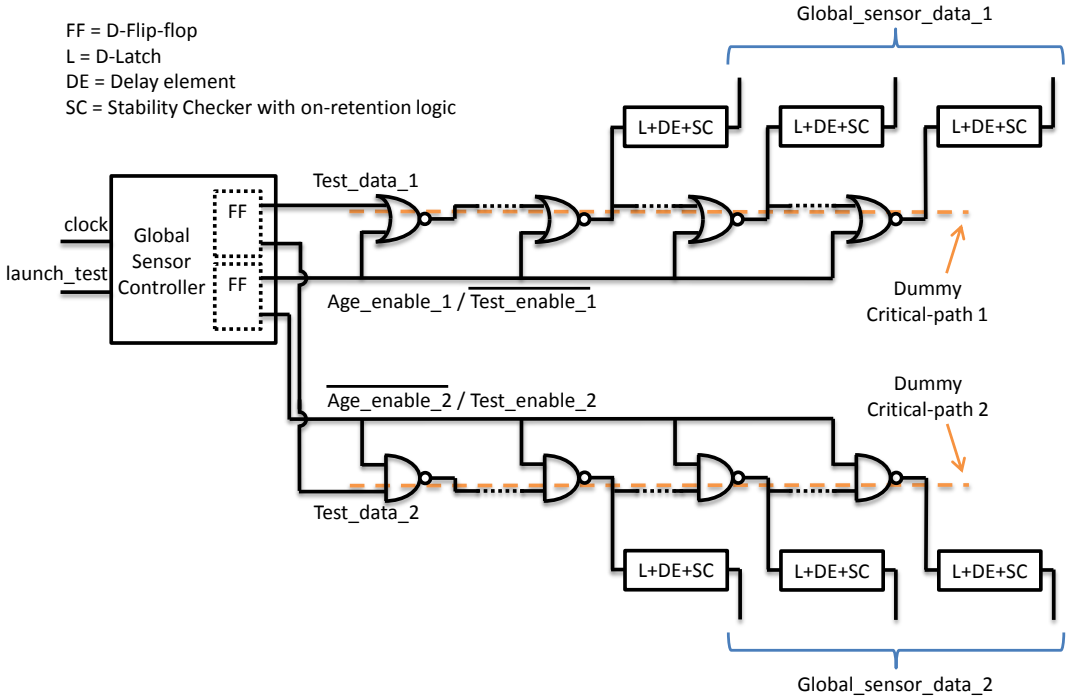


Figure 4.1 - Global Aging-Aware Performance Sensor (GAAPS) architecture.

4.1.2 NOR AND NAND GATES

The NOR input port-map is important for the high aging degradation of the PMOS transistors in the NOR chain path. Also the NAND input port-map is important for the high aging degradation of the NMOS transistors in the NAND chain path.

The internal structure of the NOR gate is presented in Figure 4.2 and the internal structure of the NAND gate is presented in Figure 4.3.

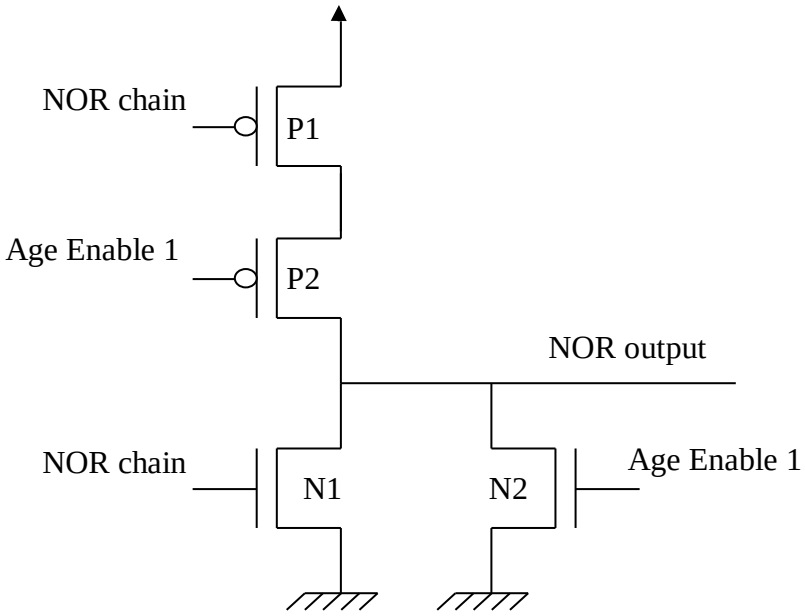


Figure 4.2 - NOR gate internal structure and port-map

For the NOR gate, the probability for the transistor P1 to be in stress mode is equal to the probability of having its NOR_chain input at logic value 0. If global sensor is activated periodically, this signal will most likely be at low logic value most of the time, making a high degradation probability for transistor P1.

However, the probability to put P2 in stress mode is equivalent to the probability of having both P1 and P2 transistors on, i.e., $P(\ 〚NOR\_chain〛\ \_input = 0) \times P(\ 〚Age\_enable\_1〛\ \_input = 0)$. Considering that global sensor is activated periodically, Age_enable_1 signal has low probability to be at 0 logic value. Henceforth, P2 will have negligible degradation. Yet, a high degradation probability of CP delay's replica is guaranteed with the high degradation probability of all the P1 transistors of the NOR chain.

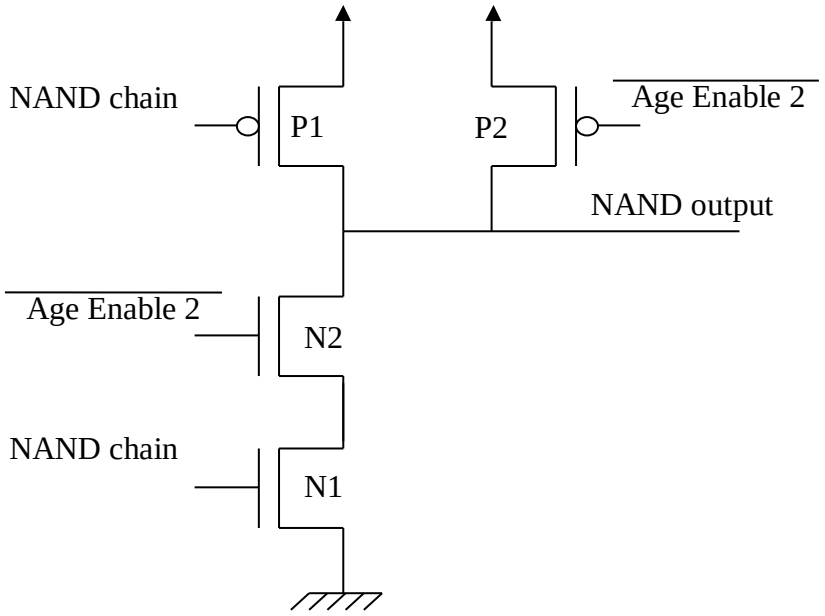


Figure 4.3 - NAND gate internal structure and port-map.

For NAND gate, the analogue of the NOR gate happens. The probability for the transistor N1 to be in stress mode is equal to the probability of having its NAND_chain input at logic value 1. If the global sensor is activated periodically, this signal will most likely be at high logic value most of the time, making a high degradation probability for transistor N1.

Though, the probability to put N2 in stress mode is equivalent to the probability of having both N1 and N2 transistors on, i.e., $P(\ [\![NAND\_chain]\!]\ \_input = 1) \times P([\![\overline{Age\_enable\_2}]\!]\ \_input = 1)$. Considering that global sensor is activated periodically, $\overline{Age\_enable\_2}$ signal has low probability to be at 1 logic value. Henceforth, N2 will have negligible degradation. Yet, a high degradation probability of CP delay's replica is guaranteed with the high degradation probability of all the N1 transistors of the NAND chain.

4.1.3 DETECTION CIRCUITRY

The detection circuitry is composed by a L+DE+SC block (D-type Latch, activated in the low state of the clock signal, a Delay Element and a Stability Checker). In fact,

this is a smaller version of the local sensor flip flop, presented in [19] and also defined in [37] and in [25] (Figure 4.4).

This block captures the critical path delay's replica. The sensor placed at the end of the CP delay's replica will be the first to flag a predictive error detection, when a PVTA variation occurs. With higher degradation in PVTA conditions, sensors with short path will flag a PVTA variation in the global sensor.

Also, the global sensor output data will allow on-line optimization of power consumption or performance through DVS or DFS respectively. An off-line tuning would allow the DVS or DFS control mechanism to update the data for which these are safe/unsafe sensors' activations, because the CP delay's replica ages differently from CUT's CPs,

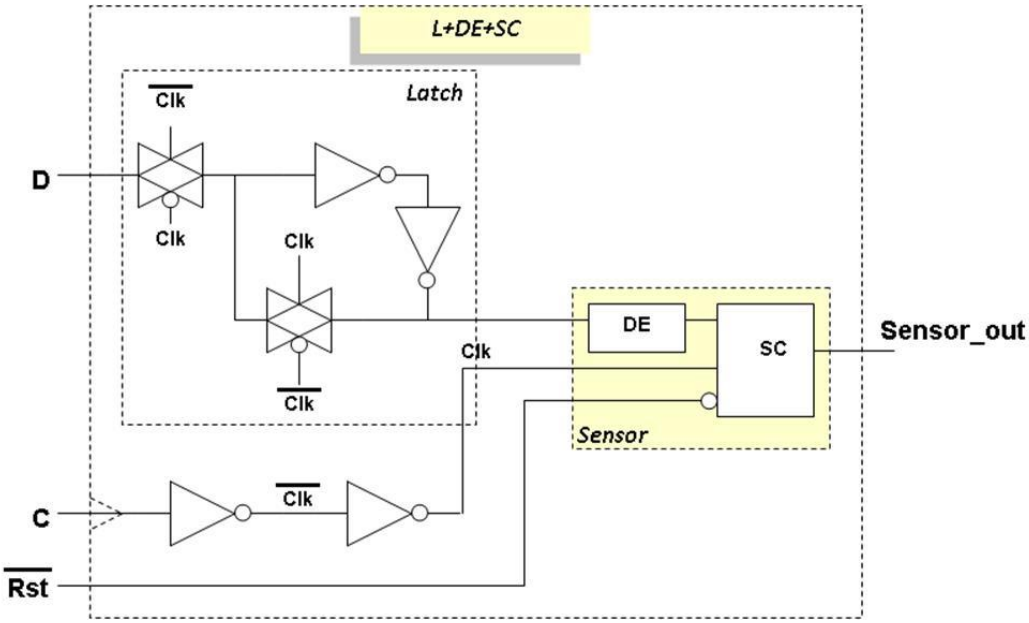


Figure 4.4 - L+DE+SC block internal structure.

4.2 LOCAL SENSOR

Local performance sensing has two purposes: (1) perform a continuous on-line monitoring of CUT's identified critical paths, by placing sensors locally at the end of each CP; and (2) perform an off-line tuning of the global sensor.

Due to PVTA variations, circuits degrade its performance over time. J. C. Vazquez [21], Agarwal [23] and C. Martins [19] proposed some local performance sensing solutions. One of those approaches is circuit failure prediction. Approaches based in circuit failure prediction seems a very promising solution and relies on aging sensors to monitor circuit performance loss.

This work is based in a previous work done by C. Martins [19] where he presented an Adaptive Error-Prediction Flip-Flop (AEP-FF) with built-in aging sensor. In the following subsections the AEP-FF architecture and functionality are explained briefly.

4.2.1 ARCHITECTURE

The topology of the proposed local aging-aware performance sensor Flip-Flop (AEP-FF) is shown in Figure 4.5.

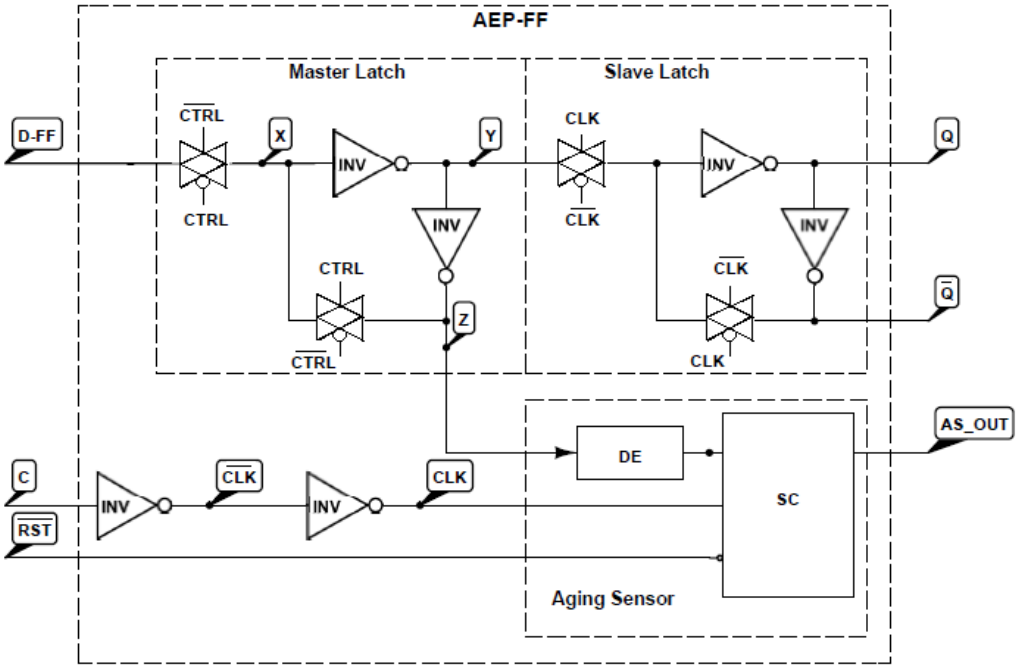


Figure 4.5 - AEP-FF Architecture.

The Delay Element (DE) delays data signals captured at the Master Latch output, during CLK low state. The Stability Checker (SC) analyzes data transitions during CLK high state. This way, the DE propagation delay is the effective observation (or

guard-band) interval. Late transitions at FF data input (propagated to the Master Latch output) will be identified by the SC. The SC has on-state retention logic, to discard the use of an additional latch to store the AS_OUT signal.

Using the flip-flop data signal at the output of the Master Latch to drive the DE, instead of the FF input data signal, D, as in previous aging sensor architectures [25], simplifies DE design. Basically, the proposed delay element is a simple buffer that introduces a delay to create a virtual guard-band where late transitions at FF data input are flagged. Furthermore, if PVTA variations occur in DE, this virtual guard-band will increase accordingly. This way, the sensor's sensitivity is adapted with PVTA variations on the DE. Another advantage is that the guard-band signal does not need to be distributed as a second balanced clock to the sensing FFs.

The loading effect of the sensor is inside the FF; hence, it does not explicitly impact the signal path [25].

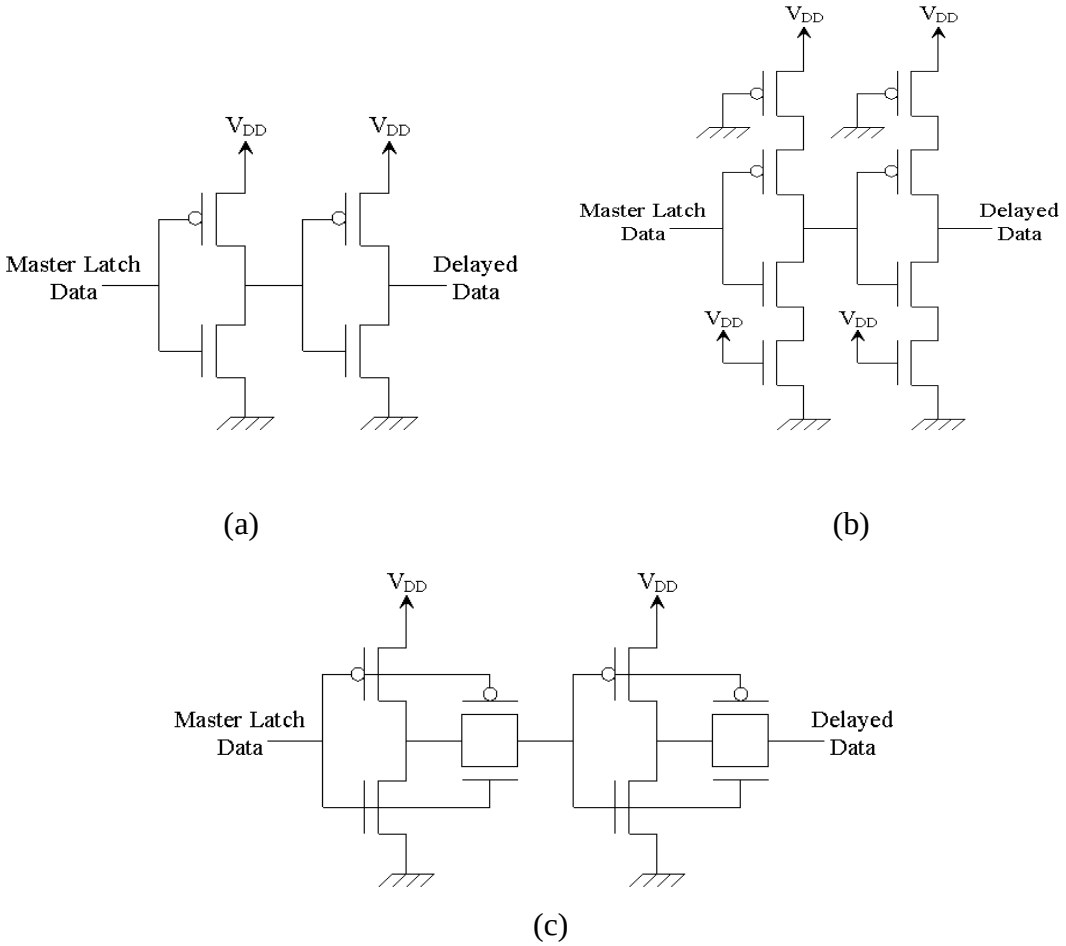


Figure 4.6 - Delay element typical architectures: (a) Low delay; (b) Medium delay; (c) High delay.

C. Martins [19] designed three DE using typical DE architectures (Figure 4.6). These buffers have different delay capability and different aging performance degradations. The factors that will influence the choice of the DE are the clock frequency, the ratio between the guard band (GB) and the clock period, the Integrated Circuits (IC) technology, and the sensor's sensitivity (or the PVTA in the SS corners for which the sensor starts to flag a late transition). The guard band is adaptive with PVTA variations, enhancing the sensor's detection sensitivity. Each sensing FF will have its own unique PVTA-dependent guard-band (each local DE may age differently).

The Stability Checker architecture is presented in Figure 4.7. The SC is implemented with dynamic CMOS logic and has built-in on-retention logic.

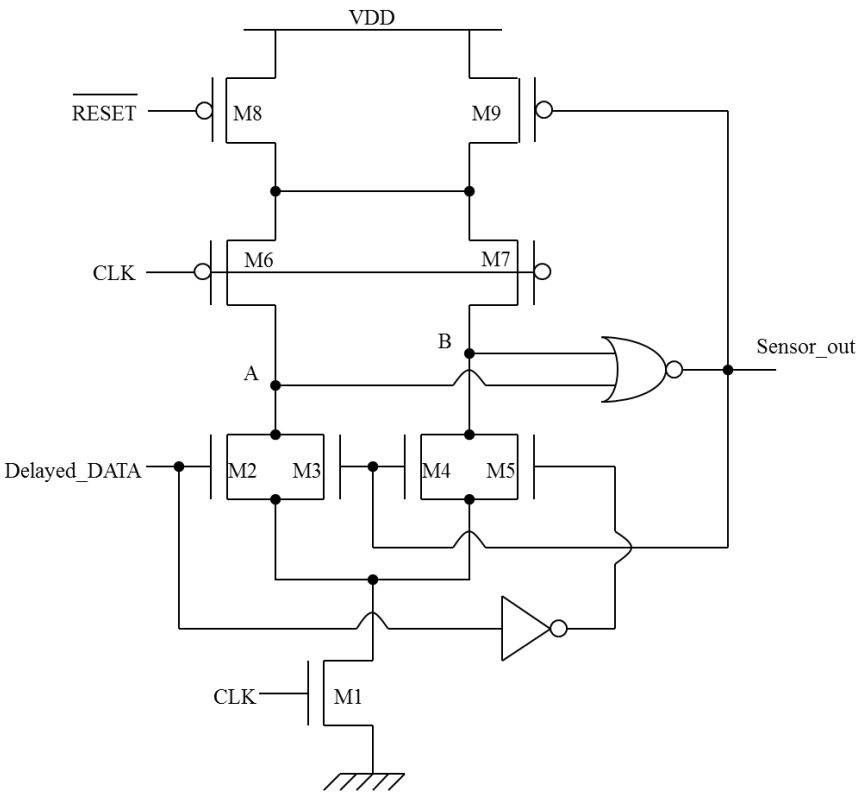


Figure 4.7 - Stability checker architecture with on-retention logic.

C. Martins [19] explains how the SC works. During CLK low state, and considering that Sensor_out signal is low, A and B nodes are pulled up (making Sensor_out to stay low). When CLK signal changes to high state, M3 and M4 are OFF, and according to Delayed_DATA signal, one of the nodes A or B changes to low. If, during the high state of the CLK, a transition in Delayed_DATA occurs, the

high A or B node is pulled down by transistor M2 or M5, respectively, driving Sensor_out to go high. From now on, M9 transistor is OFF. Hence, A and B nodes are not pulled up during CLK low state, unless the active low RESET signal is active. A and B nodes remain low, helped by transistors' M3 and M4 activation during Sensor_out high state. For the RESET signal to restore the cell's sensing capability, it must be active, at least during the low state of one clock period.

The proposed SC architecture, with the on-retention logic implemented with transistors M3, M4, M8 and M9, does not need an additional latch to retain the SC output signal when it is active. Moreover, only FF's internal clock signal is triggering the beginning of the observation interval, guard band. The end of the guard-band interval is ultimately limited by half the clock cycle (when CLK signal is high).

4.2.2 FUNCTIONALITY

A local sensor must monitor performance degradation locally, where synchronization errors may start to occur.

Thus, the aging-aware performance sensor functionality is integrated in the FF that captures CP or near-CPs. The aim is to use the AEP-FF for safety-critical, high performance systems. The purpose is to execute on-line monitoring of long-term degradation performance of the CMOS digital systems, irrelevant of their origin.

Specific flip-flops (FF) are replaced by the AEP-FF in order to monitor flip-flop data input late transitions. When the sensor signalizes the late transitions, the error prediction is made just before the error occurs, regardless of their origin e.g., caused by aging effects, or by physical defects activated by long-term operation, or even by a worst-case PVTA scenario. The AEP-FF performs locally all the monitoring procedure, so it reduces to a minimum the global interconnections for aging sensor insertion [5].

The guard-band interval is defined by design; however, its sensitivity increases with PVTA variations. So, the flip-flop sensor will adapt and increase the guard-band, as circuit variability increases with aging. This characteristic of the sensor is a key advantage, because the degradation of the sensor's performance will allow forecasting

the errors earlier. Also, the sensor is not placed in the signal path, so the performance penalty on the circuit is negligible.

The possibility of the sensors to be always active, improved the Predictive Fault-Detection (PFD), because the aging/performance degradations of the sensor itself enhances the sensor's sensibility.

The AEP-FF main features can be resumed as [19]:

- The sensor circuitry is installed locally, without additional control signals, or synchronizing signals distributed to all the sensors in a circuit;
- The identification of late transitions is made after the clock cycle's active edge. This is achieved with a delay element (DE) that delays internally the FF input data transition and creates a virtual adaptive guard-band. In the presence of performance degradations, the sensor increases its sensitivity to predict performance errors by increasing locally the virtual guard-band;
- The sensor's simplicity of usage with easy tuning and install;
- The possibility of using different DE circuits to define different guard-band depending on the application, and also the possibility to directly replace an existing critical memory element by the AEP-FF, simplifying its installation and applicability to a circuit;
- The low area overhead introduced to a circuit, compared to previous solutions;
- The low to none performance penalty on the paths being monitored, because the delayed signal is captured on the flip-flop internal nodes;
- The local sensors can be always active. These sensors are able to be active for several years without losing detection capability due to aging, a unique characteristic when compared to other aging sensors proposed before. In fact, this will increase circuit's fault-tolerance capability.

4.3 CONTROLLER AND SENSOR TUNING

The purpose of the optimization methodology is to control online DVS or DFS to optimize power consumption or performance in a circuit, while still avoiding system

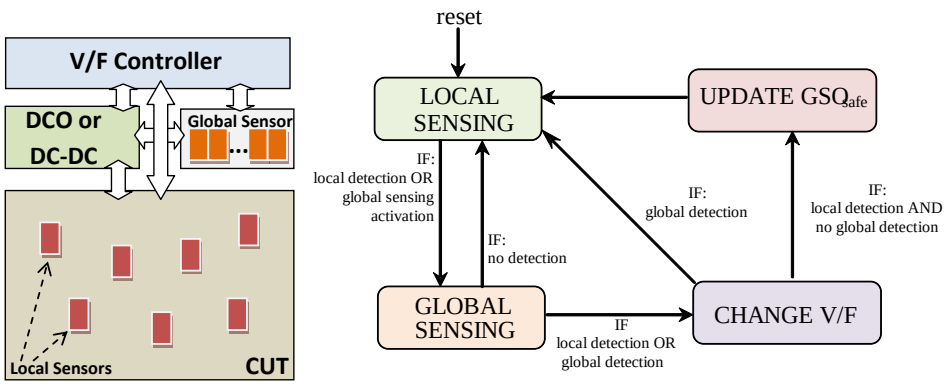
errors. The innovation is that the optimization methodology can be performed on-line and is aging-aware.

To maximize the optimization of the circuit performance, it should be initiated a sensor calibration in test mode, where some critical paths in the circuit are stimulated (using a deterministic test), and at the same time a performance analysis is made by the global sensor cells unit. This procedure, performed repeatedly for different values of the clock frequency (or voltage), will define the maximum (or minimum) frequency (or voltage) which provides a circuit operation without the local sensors being activated. This frequency information (or voltage) will be stored in a register controller (V/F register) and corresponds to the frequency (or voltage) work.

Furthermore, simultaneous operation of the sensors allows to determine which global dummy paths in both chains have similar propagation time of the circuit critical paths. This information is also stored in two registers in the system controller (GSOsafe register) and specifies the outputs of the global controllers for circuit correct (optimized) operation.

The tuning procedure can be described as follows. It starts with the highest frequency produced by the DCO (or lowest voltage produced by the DC-DC voltage regulator). Then it is applied to the CUT a set of off-line deterministic delay fault test patterns and the Local Sensor's state is analyzed. If errors occur, the sensors are initialized and the clock frequency is reduced (or the power-supply voltage is increased). This procedure is repeated for each set of tests, until no error is flagged in the Local Sensor. At the end, it can be defined, for each chip, the maximum safe frequency (or minimum safe voltage) of operation (f_{safe} or V_{DDsafe}), for which no Local Sensor was activated. Hence, it will be known the code Global Sensor output state for a safe operation, defined as a code word GSO_{safe}. The 2 words for safe operation (f_{safe} or VDD_{safe}, and GSO_{safe}) are stored in two registers, the "VF_code" and the "GSO_code" registers, respectively. These information will be used during the on-field operation.

In an on-field operation, f_{safe} (or V_{DDsafe}) will be the initial nominal frequency (or voltage), $f_{nominal} = f_{safe}$. (or $V_{DDnominal} = V_{DDsafe}$). GSO_{safe} is used to dynamically adjust frequency clock (or voltage). The Global Sensor is responsible for the coarse grain delay fault prediction, while the Local Sensors will perform a fine grain fault prediction. The global sensor activation period is defined depending on time slack. The shorter the time slack, the shorter the global sensor's activation period should be.



(a) (b)

Figure 4.8 - (a) Block diagram; (b) Simplified functionality for on-line operation of the Dynamic Voltage or Frequency Tuning controller.

Figure 4.8 depicts the block diagram of the circuit with sensors and controller, and the simplified functionality of the power or frequency optimization controller. Because local sensors are always on-line, the initial state is "Local Sensing". The state change to "Global Sensing" state is triggered by an error detected by the local sensors, or by a timer. If it is not detected by the local sensors an error, neither by the global sensors (sensor output code = stored GSO_{safe}) no action is taken, and it returns to the "Local Sensing" state. However, if a global error or a local error is detected, the state is changed to the "Change V/F" state. If a global error is detected, the frequency (or voltage) is updated (increased or reduced) and the state is changed to "Local Sensing. Though, if only local sensors detect an error, it means that at least one of the CUT's critical path ages faster than the dummy CP, and so "VF_code" and "GSO_code" registers should be updated. The frequency (or voltage) is reduced (increased) in "Change V/F" state and "GSO_code" is updated in the "Update GSO_{safe}" state with the safer global sensor output code.

If there is a possibility that aging is greater in circuit paths than the aging of the dummy CP, the existence of local sensors always monitoring the performance of the circuit ensures that the system can learn from the sensors signals and will adapt to the new operating conditions over the life of the circuit.

This optimization method and the presented topologies can be expanded and applied to other types of flip-flops, or using other types of sensors or other dummy paths in the global sensors.

4.3.1 IMPLEMENTATION

The "V/F Controller" represented in Figure 4.8 (b) was implemented in VHDL code. The simulations allow us to confirm the state transitions, the "VF_code" updates and the "GSOcode" updates (Figure 4.9).

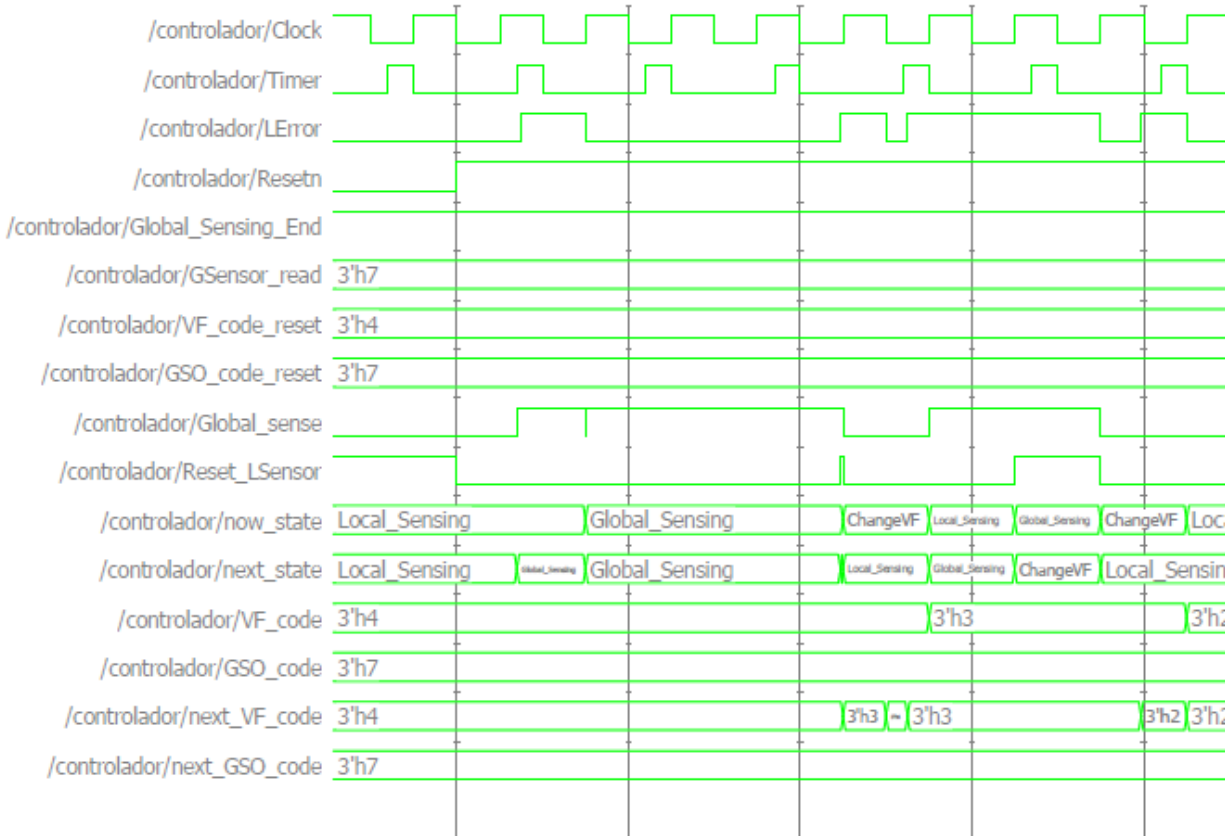


Figure 4.9 - VHDL simulation of the V/F Controller

The VHDL code was converted to Verilog code using INESC-ID software and further using AgingCalc software, developed in the University of Algarve, the Verilog code was converted in Spice Netlist.

Simulations to confirm that the "V/F Controller" was working as expected were performed in HSpice. Results of the HSpice simulations are presented in chapter 6 - Simulation Results.

4.4 DCO – DIGITAL CONTROLLED OSCILLATOR

The Digital Controlled Oscillator (DCO), was designed based on a CMOS Voltage Controlled Ring Oscillator (VCO) with Improved Frequency Stability [24].

4.4.1 THE VCO

In [24] it is shown a comparison between an invertor based oscillator and 3 types of current starved oscillator. It is shown also a comparison with combined types.

The designs given in Figure 4.10 b), c), d) are of current starved type, for which the charging and discharging output capacitor current is limited by a bias circuit.

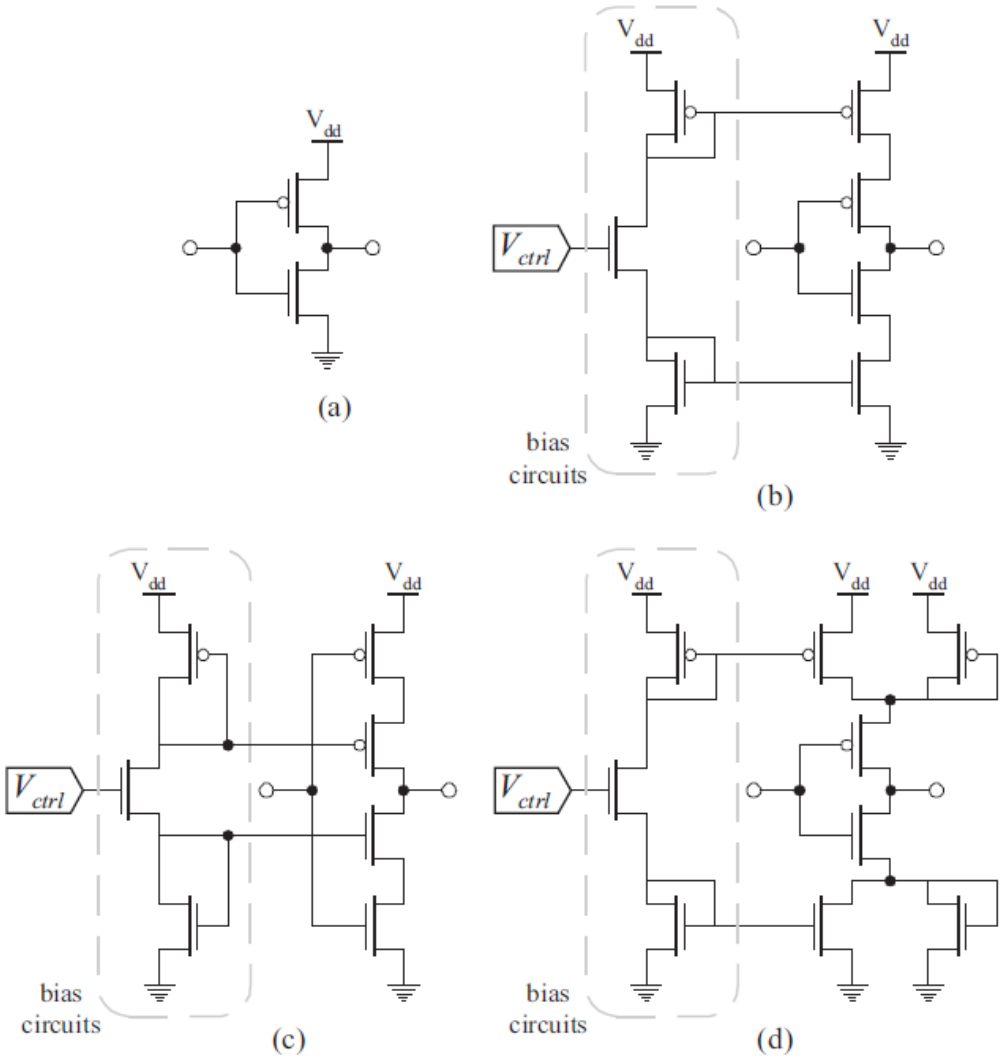


Figure 4.10 - Inverter: (a) basic type; (b) current starved with output-switching; (c) current starved with power switching; (d) current starved with symmetrical load.

4.4.2 FREQUENCY DEVIATIONS

According with [24], relative frequency deviations in terms of temperature variations, the 3-stages ring oscillators based on the inverters stages presented in Figure 4.10 have similar behavior, but the current starved with output-switching (Figure 4.10 b)) inverter has the lowest sensitivity and the ratio of relative frequency deviations between basic type (Figure 4.10 a)) and current starved with output-switching (Figure 4.10 b)) inverters is 5:1.

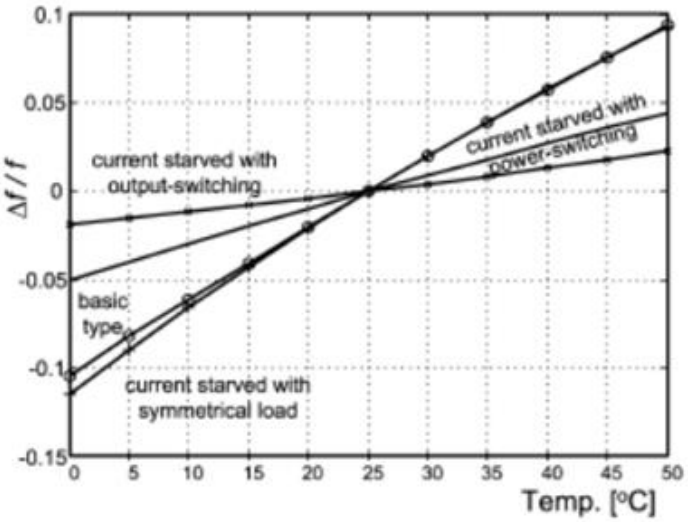


Figure 4.11 - Relative frequency deviation in term of temperature variation [24].

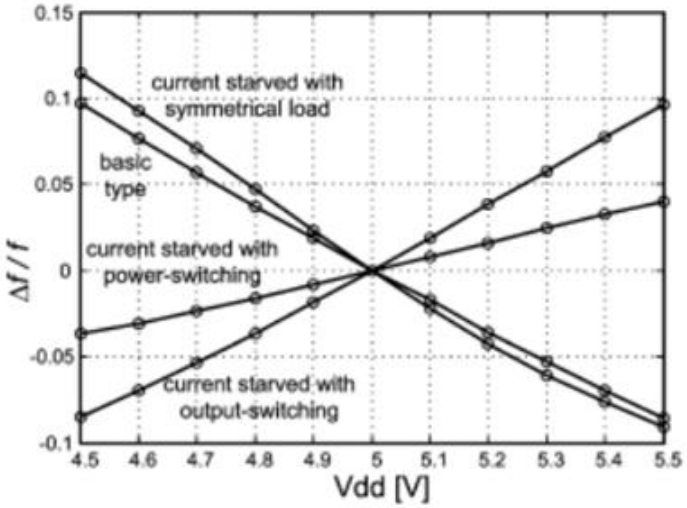


Figure 4.12 - Relative frequency deviation in term of power supply voltage variation [24].

Also, relative frequency deviations in term of power voltage supply variations for 3-stages ring oscillators based on the inverters stages presented in Figure 4.10 have similar behavior, but the current starved with power-switching (Figure 4.10 c)) inverter has the lowest sensitivity. While sensitivity in function to power supply voltage variation is within a range of 10% for all of the inverters, the current starved inverter with power-switching (Figure 4.10 c)) inverters has a lower sensitivity of 5%.

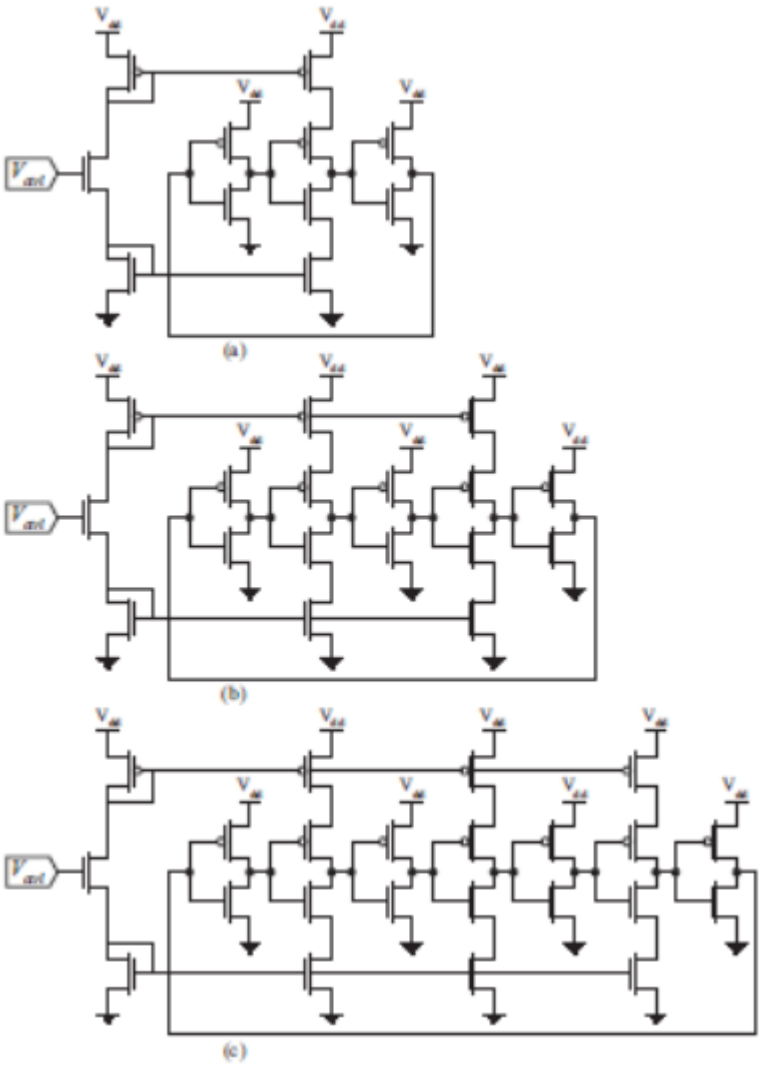


Figure 4.13 - Combined ring VCOs: (a) 3 stages; (b) 5 stages; (c) 7 stages [24].

A solution of combined ring oscillators is also presented in [24]. This type of combined ring oscillators present reduced sensitivity, and their design is done by placing basic type inverters in the odd numbered inverter stages, while even numbered are composed by current starved with output-switching inverters (Figure 4.13).

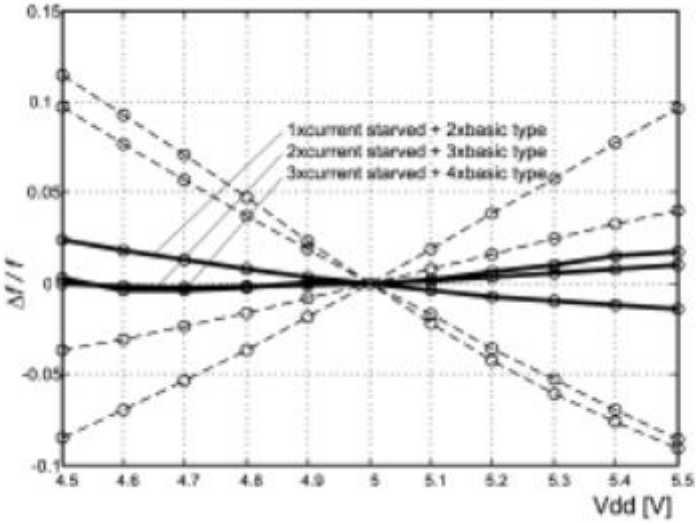


Figure 4.14 - Relative frequency deviation in term of power supply voltage variation for proposed ring VCOs [24].

The relative frequency deviations in term of power supply voltage for the combined ring oscillator's types tested are less than 2%. For the 5 stages or 7 stages oscillators, the relative frequency deviations in term of power supply voltage are less than 1%.

4.4.3 PHASE NOISE

The amount of phase noise for all types of ring oscillators (Figure 4.13) discussed in [24] varies between approximated 0.06 rad and approximated 0.3 rad.

The best phase noise performance is delivered by the ring oscillators based on current starved inverters with output switching, while the worst phase noise performance is delivered by the ring oscillators realized with basic type or current starved with power-switching inverters. Also, combined ring oscillators, have phase noise approximately within the range 0.16-0.2 rad.

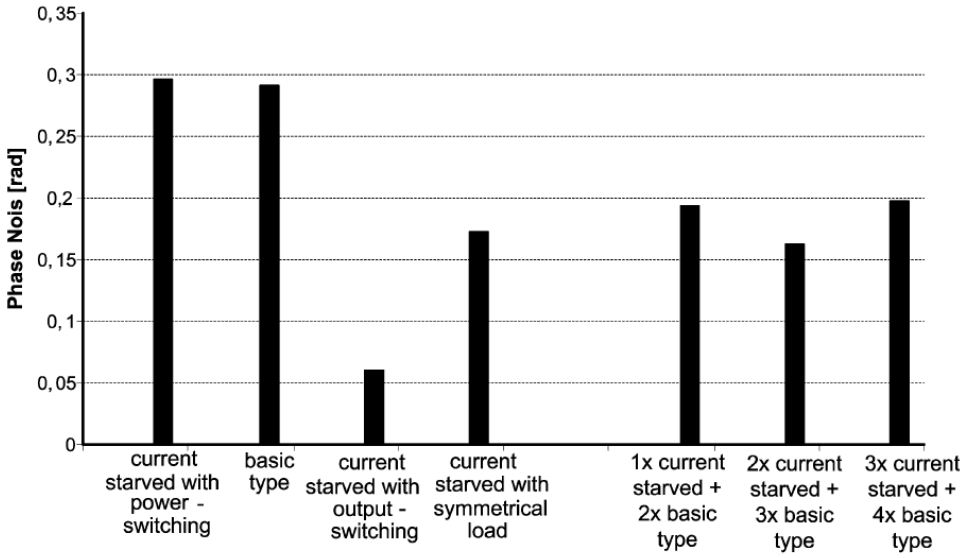


Figure 4.15 - Relative frequency deviation in term of power supply voltage variation for proposed ring VCOs [24].

4.4.4 DECISION

It was decided to use a combined ring oscillator with 5 stages (2 X Current Starved + 3 X Basic Types) because:

- for frequency stability in terms of power supply voltage variations, combined types of ring oscillator with 5 or 7 stages, presents the best performance deviation (less than 1%). This is an improvement compared to the standard solutions where the best performance is 4%.
- In respect to phase noise, the 5 stages combined ring oscillator is the second best (approximately 0.16 rad) only exceeded by the current starved with output-switching (approximately 0.06 rad) but with worst frequency deviation in terms of power supply voltage variations (approximately 10%).

4.4.5 THE DCO

It was needed a DCO with 8 different frequencies, varying between 1GHz and 3GHz, and controlled by a 3-word data.

So we used the combined ring oscillator with 5 stages, and connect to the bias circuit a voltage divisor made with CMOS.

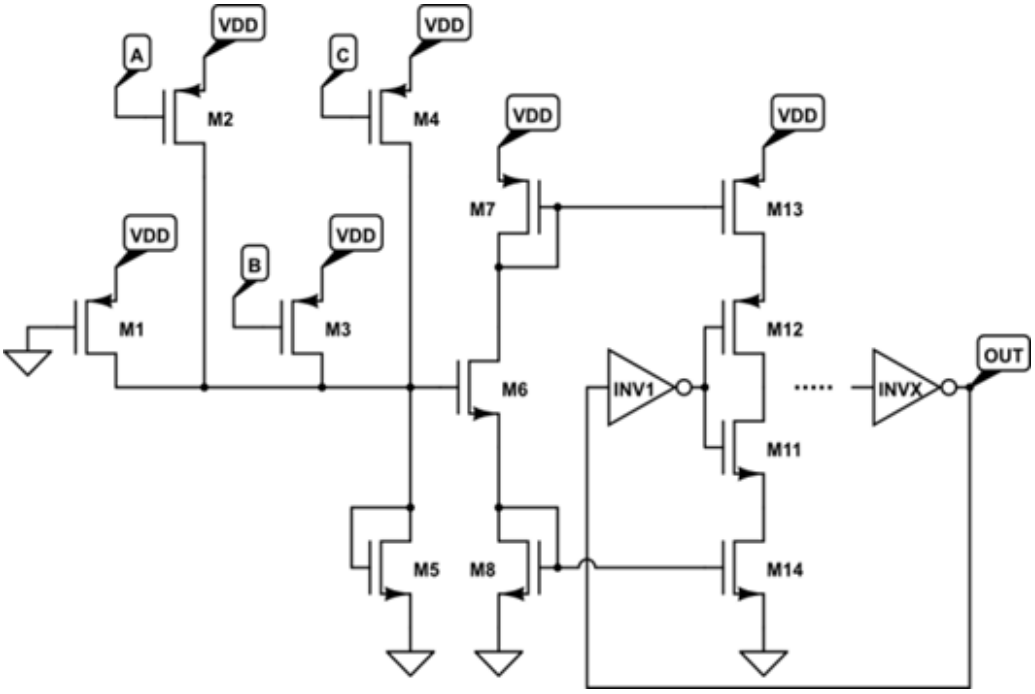


Figure 4.16 - DCO architecture.

In order to have different voltages in the Vctrl to achieve different frequencies, the sizes of the M2, M3 and M4 must be different between themselves. The width (W) of the M2, M3 and M4 is 2, 4 and 8 times bigger of the minimum size of the technology used (65nm) respectively. The size of the M1 ($W=W_{Pmin}$ X 3.25) and M5 ($W=W_{Nmin}$ X 6.25) was set in order to have the voltage between 0,69V and 0,975V in the V_{ctrl}. These voltages allow reaching the range of frequencies desired (1GHz-3GHz).

Figure 4.16 presents the DCO structure and Figure 4.17 presents the frequencies for the 8 ABC input combinations. Therefore a 3-word data controls the 8 possible frequencies for the DCO.

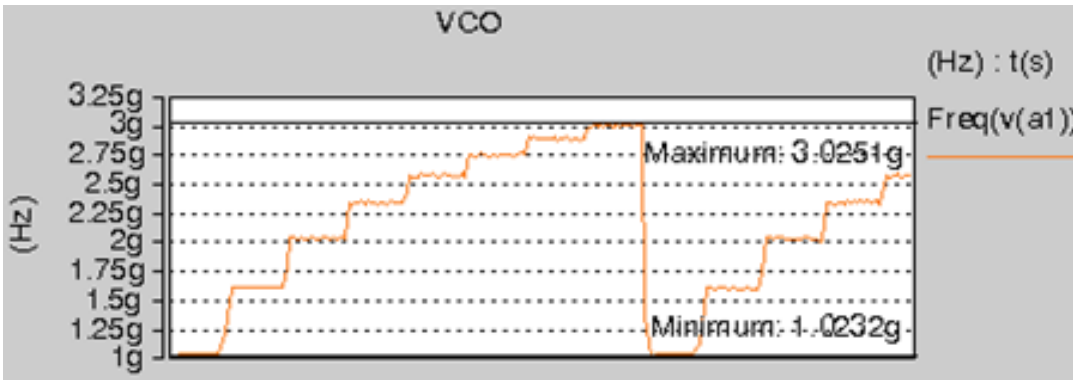


Figure 4.17 - DCO frequencies for the 8 ABC combinations.

5. ADAPTIVE ERROR PREDICTION AND DELAY–FAULT CORRECTION FLIP-FLOP

The present chapter presents the main characteristics of the DFC-FF (Delay-Fault Correction Flip-Flop) and of the Adaptive Error Prediction and Delay–Fault Correction Flip-Flop (AEPDFC-FF) developed in this thesis work, and based in the AEP-FF (Adaptive Error-Prediction Flip-Flop) aging sensor developed in C. Martins [19] thesis.

5.1 CONCEPT

The purpose of this new sensor is to correct late transients in the data input of the flip-flop (FF). This functionality is an addiction to the previous AEP-FF developed in C. Martins [19] MSc Thesis, however it can be used in the a flip-flop without the aging sensor capability.

The AEP-FF is a sensor design with a virtual guard-band time defined by the delay time propagation of a delay element (DE). The DE is placed inside the FF and its purpose is to delay the signal data at the end of the critical path.

The clock period (T_{CLK}) is greater than $\tau_0 + \tau_{GB}$ [τ_0 (critical and near-critical paths delay) and τ_{GB} (virtual guard-band)] under normal operation conditions. However, aging due to PVTA variations will increase delays (τ_0 and τ_{GB}) decreasing the time slack (τ_{slack}). This means that when $\tau_0 + \tau_{GB}$ exceeds the clock period [worst-case operation condition (WCC)] an unsafe propagation delay time is detected (though this is not enough to induce an error). The error prediction is executed during the clock positive semi-cycle (Figure 5.1) [19].

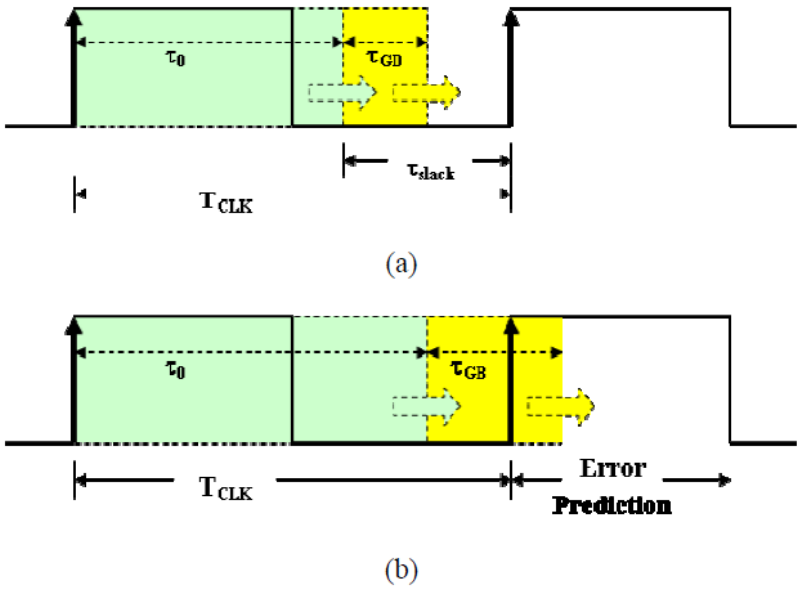


Figure 5.1 - Error-prediction and sensor operation. (a) Nominal PVTA conditions, with no error predicted; (b) PVTA WCC and error prediction.

The purpose of the DFC-FF/AEPDFC-FF sensor is to detect and correct late transients in the data input of FF that terminate critical and near-critical paths (Figure 5.2). This means that when $\tau_0 \geq T_{CLK}$, the DFC-FF/AEPDFC-FF must detect and correct the error. The theoretic maximum delay admitted by the DFC-FF/AEPDFC-FF is $\tau_0 \leq \left(T_{CLK} + \frac{T_{CLK}}{2}\right)$. However the detection window should not be larger than $T_{CLK} + \frac{T_{CLK}}{4}$ because it is an adaptive window. This means that the PVTA degradations that will induce increased delays also will increase the detection and correction window size. So during design, detection window must be defined considering that aging will affect this window.

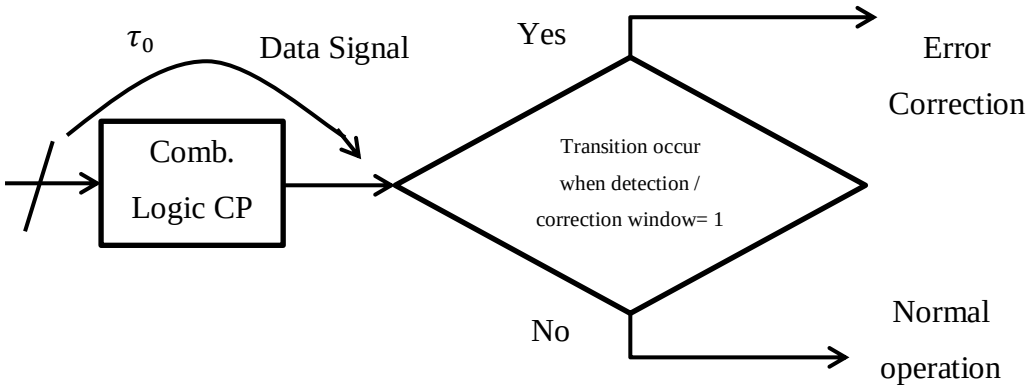


Figure 5.2 - DFC-FF concept block diagram.

5.2 TOPOLOGY

The topology of the proposed Adaptive Error Prediction and Delay–Fault Correction Flip-Flop (AEPDFC-FF) is shown in Figure 5.3, while the topology of the proposed Delay-Fault Correction Flip-Flop (DFC-FF) is shown in Figure 5.4. The AEPDFC-FF is based in the AEP-FF with an additional Correction Error Sensor (CES) [see Figure 5.5] while the DFC-FF is based in the AEPDFC-FF without the Aging Sensor.

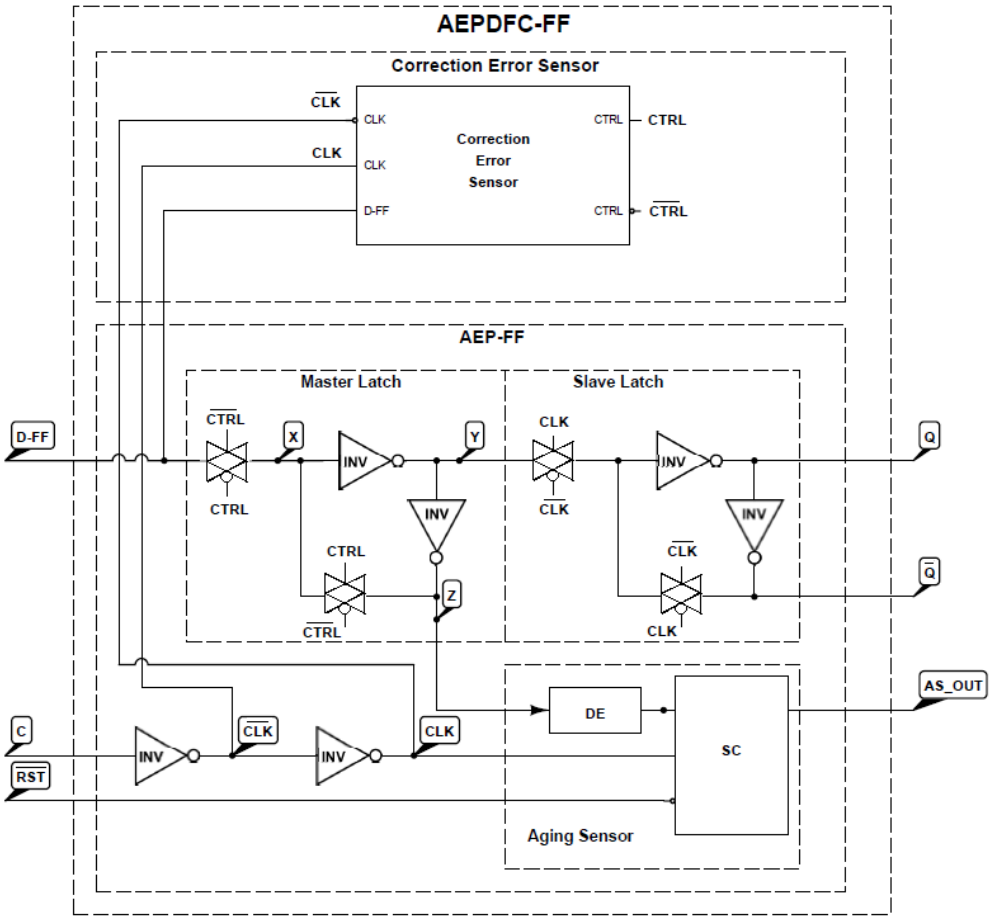


Figure 5.3 - AEPDFC-FF Architecture.

The CES will first create a detection window (DW) in the beginning of the CLK high state. This is done with a NOR of the CLK signal inverted with a delayed CLK signal. The DW size will be defined by the delay of the delayed CLK signal. This delay is obtained with a Delay Element (DE-CES).

Then the CES will verify if during the high state of the DW, there is a transition of the D signal of the flip-flop. For this it will use a Stability Checker (SC-CES) without retention logic. Because it is necessary to eliminate the setup and hold time of the flip-flop, in order to not have a failed detection or correction, a Delay Element (DE_D-CES) is added to the D signal of the flip-flop. Of course, this will cost a reduction of the DW with the same value of the delay added to the D signal.

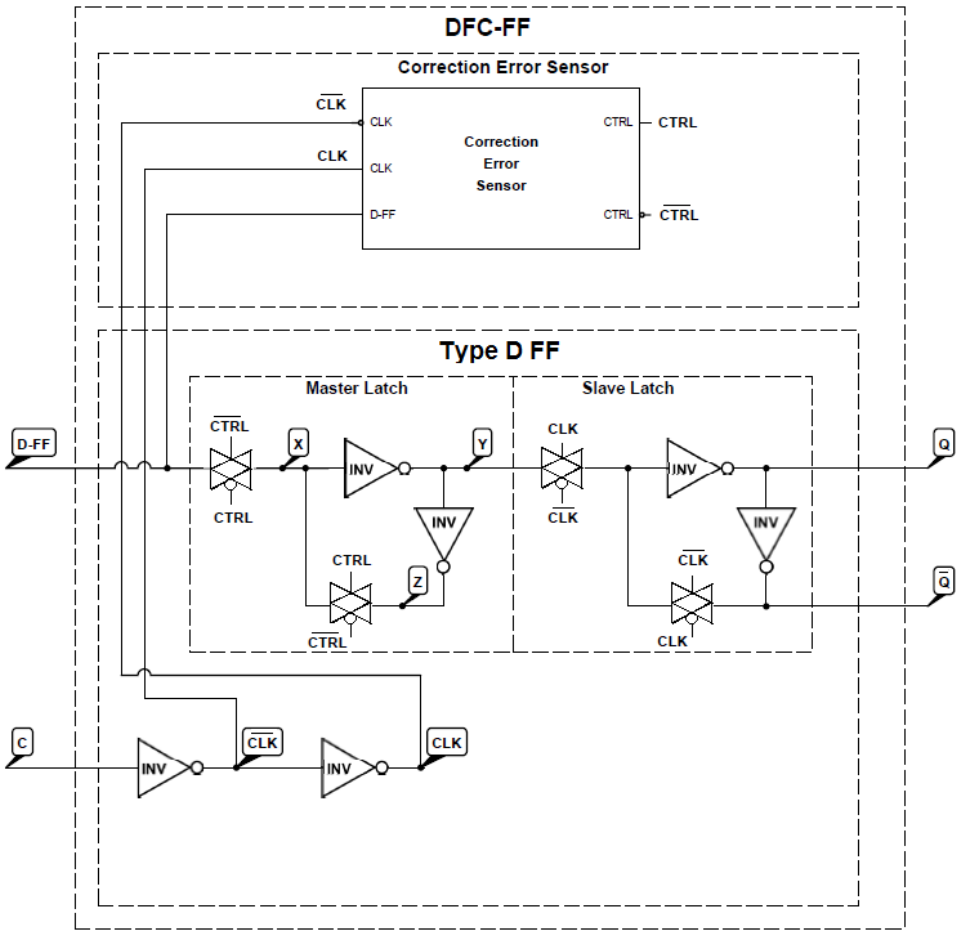


Figure 5.4 - DFC-FF Architecture.

If a late transition is detected, the FF Master-Latch must became transparent in order to correct the late transition detected. That is achieved by controlling the FF Master-Latch transmission gates with a NOR of the inverted CLK ($\overline{CLK}$) signal with the result of the SC-CES, instead of the CLK signal (Figure 5.5).

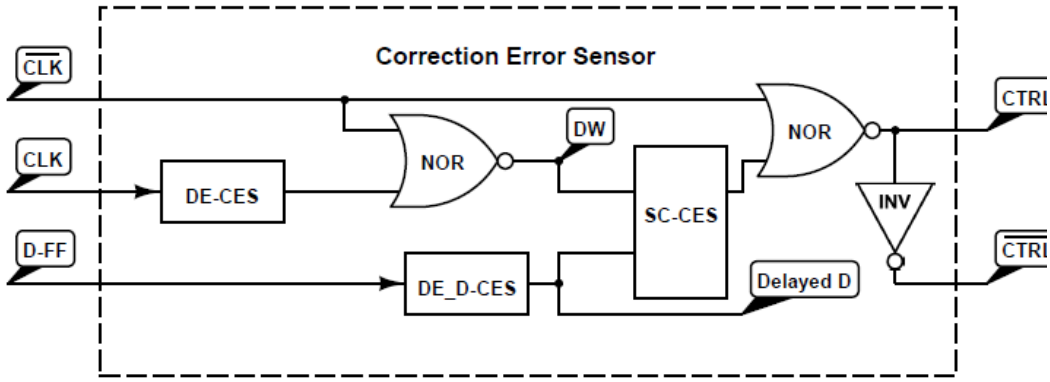


Figure 5.5 - Correction Error Sensor Architecture.

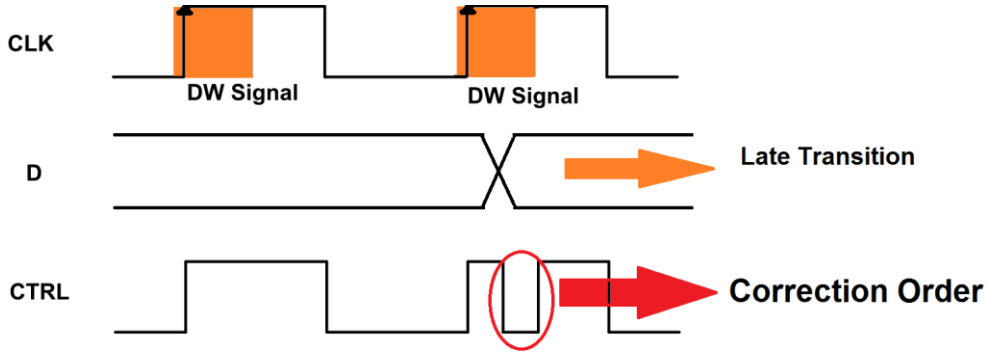


Figure 5.6 - Correction Error Sensor Diagram.

Also, if PVTA variations occur in DE-CES, the DW will increase accordingly. So, the sensor's sensitivity is adapted with circuit's cumulative aging degradation.

5.3 CIRCUIT MODULES

The DFC-FF is composed by the CES and by the AEP-FF without the Aging Sensor (a common FF type D). The AEPDFC-FF is composed by the CES and by the AEP-FF. This section will explain only the CES circuit modules while the AEP-FF circuit modules are explained in C. Martins [19].

5.3.1 DELAY ELEMENTS (DE-CES AND DE_D-CES)

Two Delay Elements are needed to design the CES. The DE_D-CES and the DE-CES.

The DE_D-CES function is to delay the D signal of the flip-flop in order to cover the response time of the CES. The signal delayed will be the signal entering the SC-CES.

The circuit architecture proposed for the DE_D-CES module is composed just by inverters connected in series. Because the SC-CES will only evaluate if there is a variation in the signal (is not important if the signal is high or low), the DE_D-CES may be composed by an odd or even number of inverters depending on time needed to cover the response time.

The DE-CES function is to define the width of the Detection Window (DW). The larger the delay introduced by the DE-CES, the larger will be the DW, with the maximum of half clock cycle (when the clock cycle is high). However, typically, the DW will be defined as less than a ¼ of the period.

In this work, it was decided to use the delay elements developed for the AEP-FF described in [19]. In order to have a larger DW, it can be connected in series several Delay Elements or a chain of inverters with an even number of inverters.

It is also interesting to note that the delay added by the DE_D-CES will cost a reduction in the DW, reducing its detection width. This means that if the DE_D-CES add “X”ps, the DW will be equal to DW minus “X”ps. So, during design this situation must be considered. Also, the delay added by DE_D-CES is exposed to PVTA variations that will lead to the increase of the delay, reducing the DW. However, also the DE-CES is exposed to PVTA variations, leading to an increase of the DW. Both effects are opposite.

5.3.2 STABILITY CHECKER (SC-CES)

The Stability Checker (SC-CES) architecture is showed in Figure 5.7. The architecture of the SC-CES was achieved by simplifying the Stability Checkers presented by [19] and [23]. The SC should be resilient to aging but did not need to

have the on-retention logic. The purpose of the SC-CES is to give the correction instruction in case a delay error occurs and not to inform about the error, so on-retention logic is unnecessary. Also, this by itself will reduce the size of the SC-CES, reducing the area overhead needed for correction.

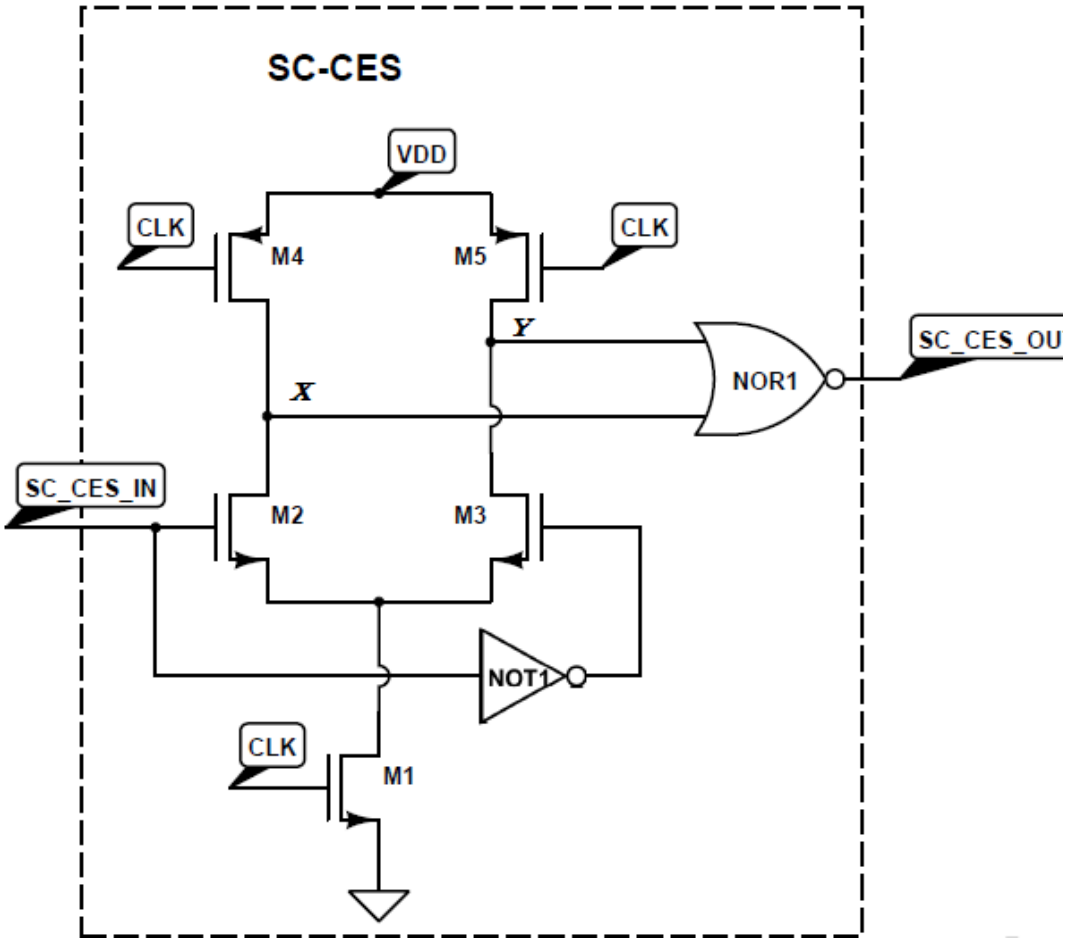


Figure 5.7 - Stability Checker Architecture.

The SC-CES will mark an error when the SC_CES_OUT is on high state. For this to happen several steps are necessary and is only possible due to high impedance property of the M4 and M5 transistors. When CLK signal is on low state the nodes X and Y are on high state. The change of the CLK signal to high state will turn the M4 and M5 transistors to OFF and according to SC_CES_IN signal, one of the nodes X or Y changes to low. If a transition in SC_CES_IN signal occurs during the high state of the CLK, the high X or Y node is pulled down by the transistor M2 or M3, respectively driving SC_CES_OUT signal to go high. When the CLK signal changes to low state, X and Y nodes are pulled up and SC_CES_OUT changes to low.

The Detection Window (DW) is the clock signal for the SC-CES that will trigger the beginning of the observation interval, and is typically a quarter of the period (when clock cycle is high).

5.3.3 CORRECTION ERROR SENSOR (CES)

The Correction Error Sensor (Figure 5.5) will first create a detection window (DW) in the beginning of the CLK high state. Then will verify if during the high state of the DW, there is a transition of the D signal of the flip-flop.

If a late transition is detected, the FF Master-Latch is changed to transparent mode in order to correct the late transition detected. However, if the DW grows and becomes larger than the positive cycle of the clock, the CES should stop giving the correction order. That is achieved with the NOR gate of the inverted CLK ($\overline{CLK}$) signal with the result of the SC-CES resulting in the CTRL signal.

However, if needed, the final NOR gate of the CES can be replaces by a XOR gate, and the CES will continue to detect and giving correction information in both cycles of the CLK. Another concern is to assure that there is no Short-Path (SP) that will arrive during DW monitoring, becoming a false error correction (as described in [8]).

During simulations the limits of this block were reached. For frequencies higher than 2GHz the response time of the CES becomes slow and starts to fail detections.

For the situations where higher frequencies are required, an alternative CES architecture is presented Figure 5.8, a Fast-CES.

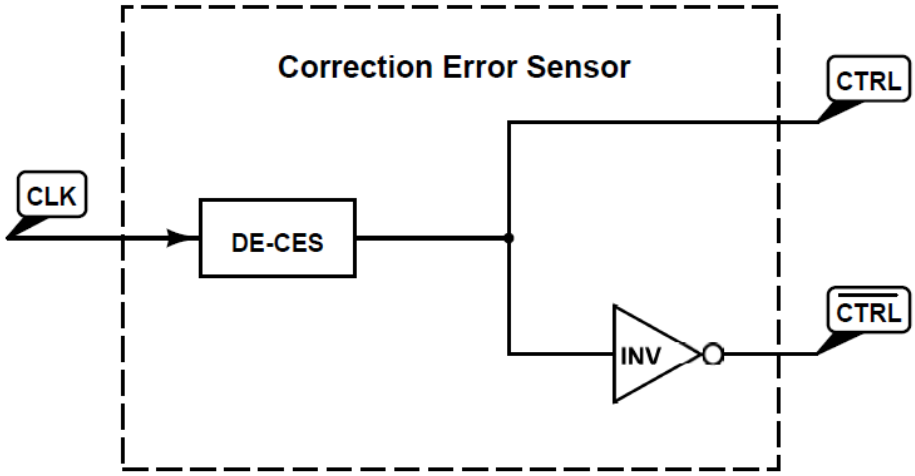


Figure 5.8 – Alternative Correction Error Sensor Architecture (Fast-CES).

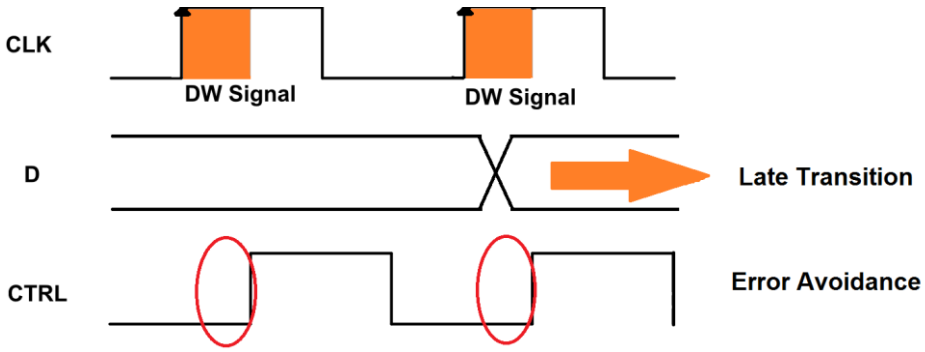


Figure 5.9 – Alternative Correction Error Sensor Diagram (Fast-CES).

The Fast-CES will delay the CLK signal creating a window of detection where transitions will be detected. This element can replace the regular CES connecting only the CLK pin and using the CTRL and $\overline{CTRL}$ to transmission gates of the first latch of the Flip-Flop. The delay introduced by the DE-CES will create a time where both latches are connected simultaneously, and a time where both are disconnected. The main difference between the regular CES and the Fast-CES is that the first will only correct if an error is detected while the Fast-CES will always have a transparent period and though not detecting any error.

5.4 CHARACTERIZATION

HSPICE software was used to characterize the DFC-FF and the AEPDFC-FF, through several simulations. The characteristics of the DFC-FF and AEPDFC-FF are addressed in the following subsections.

Simulations were performed in order to characterize the DFC-FF and the AEPDFC-FF for power-supply and temperature design corners, and to determine the delay introduced by the regular CES, when compared with the regular flip-flop and the AEP-FF.

The measurements were performed under the circumstances expressed in Table 5.1. V_{DD} variation is ±10% for FF (Fast-Fast) or SS (Slow-Slow) from NC (Normal Conditions) V_{DD} and for temperature ambient variation were chose the limits expressed in Table 3.1.

	FF +10%	**NC**	**SS -10%**
TECHNOLOGY MODEL (ηm)	PTM 65		
FREQUENCY (GHz)	1		
VDD (V)	1.21	1.1	0.99
TEMPERATURE (ºC)	-55	27	125

Table 5.1 - Design Corners for Power-Supply and Temperature values.

Simulations used a benchmark circuit, the same used to demonstrate the AEP-FF applicability [19]. It was selected a circuit composed by an inverter chain (20 inverters), due to its simplicity and to compare results with the AEP-FF.

5.4.1 PROPAGATION DELAY ANALYSIS

The propagation delay of the AEP-FF is approximately the same of the flip-flop type D [19]. In order to study the impact of the introduction of the CES in the flip-flop, it was measured the propagation delay introduced by the new flip-flops [Delay–

Fault Correction Flip-Flop (DFC-FF) and Adaptive Error Prediction and Delay–Fault Correction Flip-Flop (AEPDFC-FF)]. The circuits used are presented in Figure 5.10.

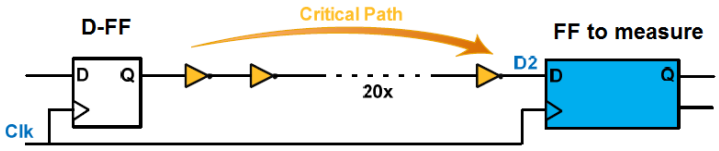


Figure 5.10 - Schematic of circuits used to measure propagation delay with type D flip-flop, with AEP-FF, with DFC-FF and with AEPDFC-FF.

The circuit composed by an inverter chain of 20 inverters is used to simulate a critical path (CP). The propagation delay is measured between the CLK signal and the signal in D2, as is showed in Figure 5.11.

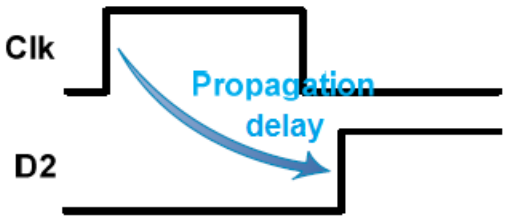
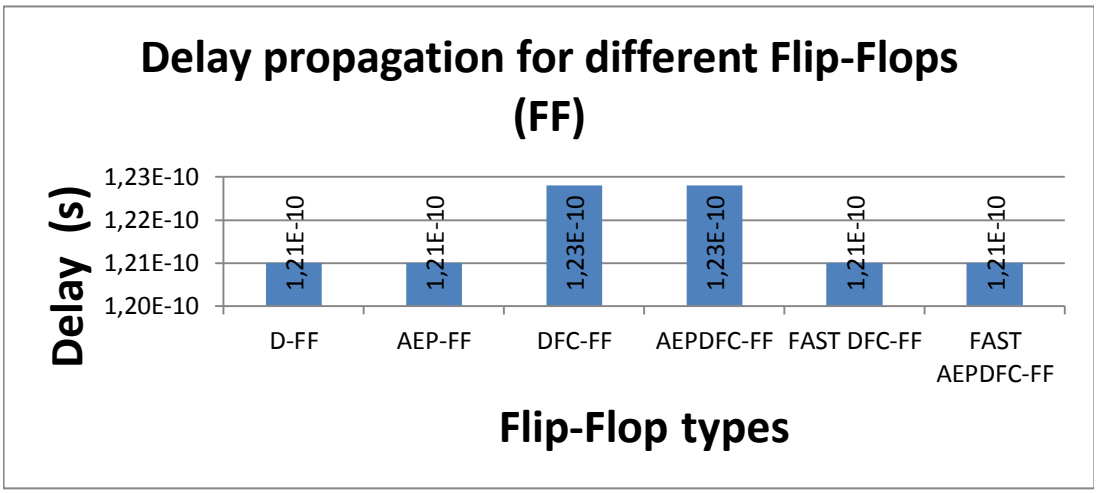


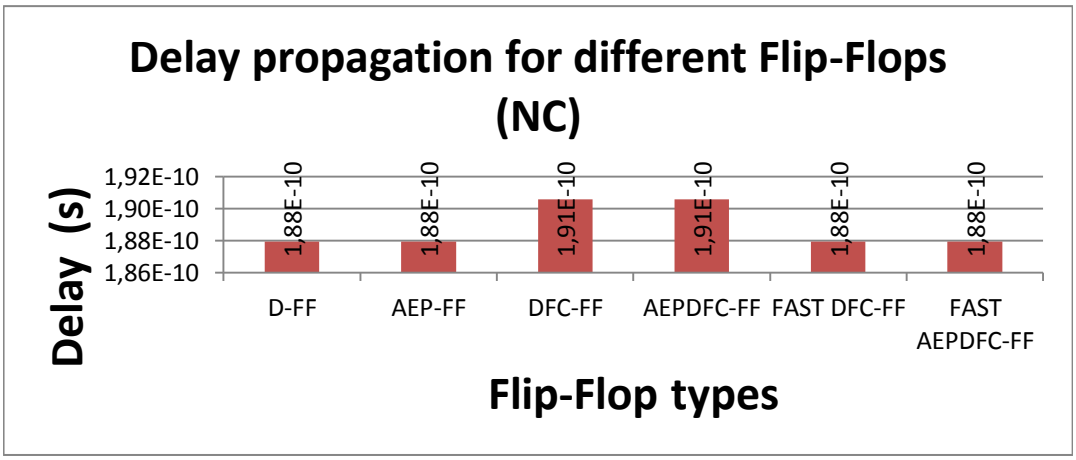
Figure 5.11 - Propagation delay definition.

Delays in ps		FLIP-FLOP types					
		D-FF	AEP-FF	DFC-FF	AEPDFC-FF	FAST DFC-FF	FAST AEPDFC-FF
Working Conditions	FF	121,00	121,00	122,80	122,80	121,00	121,00
	NC	187,95	187,95	190,60	190,60	187,95	187,95
	SS	333,95	333,95	338,50	338,50	333,95	333,95

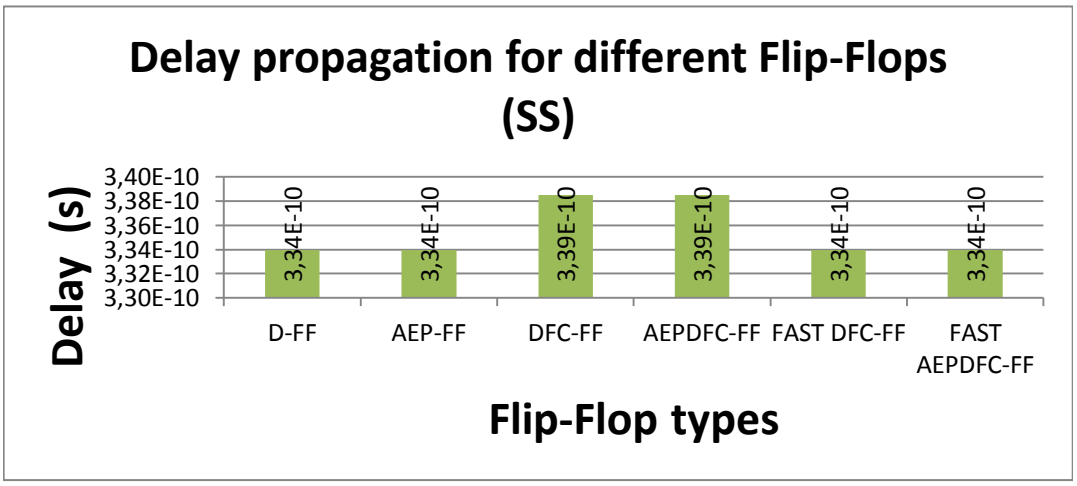
Table 5.2 - Propagation delay measured values in ps.



Graphic 5.1 - Propagation delay measured values for different flip-flops and FF conditions.



Graphic 5.2 - Propagation delay measured values for different flip-flops and NC conditions.



Graphic 5.3 - Propagation delay measured values for different flip-flops and SS conditions..

While the AEP-FF propagation delay is not affected, the DFC-FF and the AEPDFC-FF propagation delay are affected when compared with the propagation delay of a type D flip-flop. However, the results show that there is only a minor impact. The data measured is presented in Table 5.2, in Graphic 5.1, in Graphic 5.2 and in Graphic 5.3.

The same simulations but replacing the regular CES with the Fast-CES in the DFC-FF and in the AEPDFC-FF flip-flops, shows that the propagation delay is not affected, neither in the DFC-FF neither in the AEPDFC-FF. This is explained because Fast-CES doesn't make any connection in the main path of the flip-flop, thus not affecting the propagation delay.

5.4.2 DETECTION WINDOW ANALYSIS

The Detection Window (DW) theoretically is a guard band where, if the flip-flop fails, it allows to recover from that failure.

The DW width is defined by design and depends on clock frequency. The delay element introduced in the CES that will create the DW is defined in order that for

normal conditions the DW width will be smaller than ¼ of the clock frequency. However, this can be defined differently.

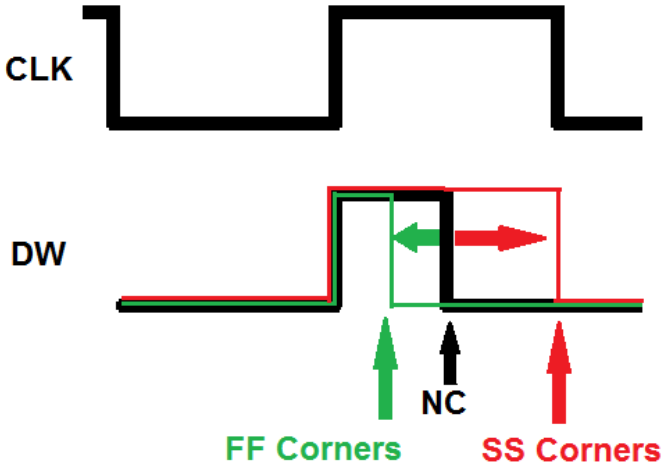


Figure 5.12 - DW width definition.

The DW is adaptive because it will vary with the operation corners. In situations that the circuit operates slower, the DW will grow, and in situations that the circuit operates quicker, the DW it will decrease.

Another characteristic is that the measured DW is smaller than the theoretical DW. See Figure 5.13.

For all operating conditions (SS, NC and FF) the DW width was measured and compared with the theoretical DW width. The conclusion is that the measured DW is approximately 70% of the theoretical DW (Table 5.3 and Table 5.4).

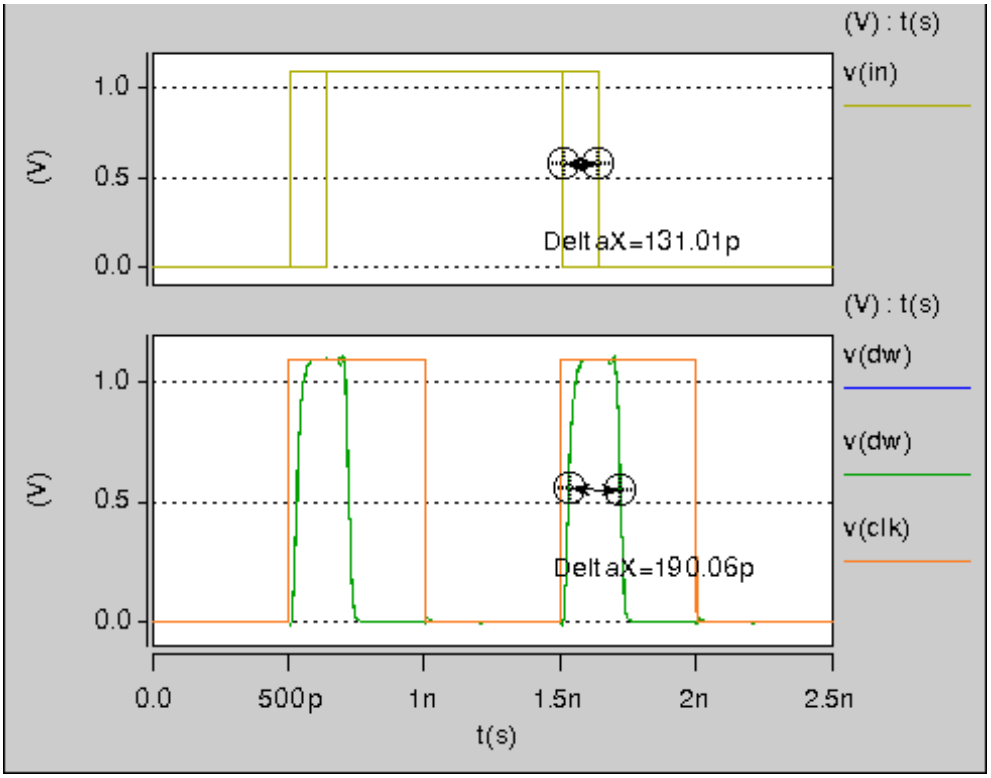


Figure 5.13 - Measured DW vs Theoretical DW.

	FF		NC		SS	
	LH (ps)	HL (ps)	LH (ps)	HL (ps)	LH (ps)	HL (ps)
Theoretical DW (ps)	114,7	114,0	192,4	190,5	367,9	363,9
Measured DW(ps)	77,0	79,0	137,0	131,0	270,0	249,0
% of Measured DW (%)	67,1%	69,3%	71,2%	68,8%	73,4%	68,4%

Table 5.3 - Detailed comparison between Effective DW and Measured DW for DFC-FF and AEPDFC-FF with regular CES.

	FF	NC	SS
Theoretical DW (ps)	114,4	191,4	365,9
Measured DW(ps)	78,0	134,0	259,5
% of Measured DW (%)	68,2%	70,0%	70,9%

Table 5.4 - Summarized comparison between Effective DW and Measured DW for DFC-FF and AEPDFC-FF with regular CES.

Also, simulations replacing the regular CES with the Fast-CES in the DFC-FF and in the AEPDFC-FF flip-flops, shows that the measured detection window is approximately 76% of the theoretical DW for the DFC-FF with Fast-CES and

approximately 82% of the theoretical DW for the AEPDFC-FF with Fast-CES. (Table 5.5 and Table 5.6)

	FF	NC	SS
Theoretical DW (ps)	48,0	95,0	181,5
Measured DW(ps)	36,0	69,0	144,0
% of Measured DW (%)	75,0%	72,6%	79,3%

Table 5.5 - Summarized comparison between Effective DW and Measured DW for DFC-FF with Fast-CES.

	FF	NC	SS
Theoretical DW (ps)	48,0	95,0	182,0
Measured DW(ps)	38,0	75,0	158,0
% of Measured DW (%)	79,2%	78,9%	86,8%

Table 5.6 - Summarized comparison between Effective DW and Measured DW for AEPDFC-FF with Fast-CES

This results, shows that if there is no need to signal the error, the Fast-CES is a better solution, for its reduced complexity and for its higher performance.

6. SIMULATION RESULTS

The simulation results presented in this chapter were obtained using a single circuit under test (CUT). The purpose of these simulations is to demonstrate the applicability of the DFC-FF and AEPDFC-FF, and also to demonstrate that the Power or Frequency control mechanism of the DVFS methodology developed can effectively optimize on-line the performance, by increasing operating frequency to the maximum limit that prevents errors' occurrence, or to restrict power consumption, by reducing power-supply voltage to the minimum value that prevents errors from happening.

The circuit used is a pipeline multiplier with 4 input bits, 2 pipeline stages, with a 8-bit output. It has 52 logic gates, 10 PI (Primary Inputs), 8 PO (Primary Outputs) and 36 FF. It demonstrates how several sensors in the same circuit are used to monitor a complex circuit, detect delay degradation at different moments in time (for the AEPDFC-FF) and correct delay errors when they start to occur during a limited correction window (for the DFC-FF and AEPDFC-FF). The AEP-FF sensor characteristics are also compared with the DFC-FF and AEPDFC-FF sensors characteristics.

6.1 AEPDFC-FF AND DFC-FF

6.1.1 SIMULATIONS CONDITIONS

The measurements were performed under the SS conditions. Table 6.1 has a resume of the values used for the simulations. HSpice simulations were performed with a sweep of the clock signal frequency in order to detect the period where the Flip-Flop start to fail.

	CONDITIONS
TECHNOLOGY MODEL (ηm)	PTM 65
PERIOD RANGE (ps)	250 to 400
FREQUENCY RANGE (GHz)	2.5 to 4
VDD (V)	0.99
TEMPERATURE (ºC)	125

Table 6.1 – Simulation conditions.

The 2-stage, 4-bit pipeline multiplier (PM) used as benchmark circuit is represented in Figure 6.1.

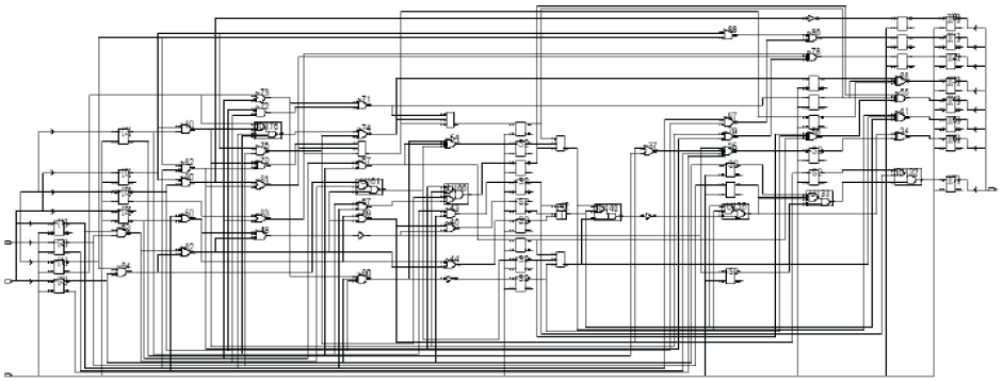


Figure 6.1 - 2-stage, 4-bit pipeline multiplier.

6.1.2 MAXIMUM CLOCK FREQUENCY

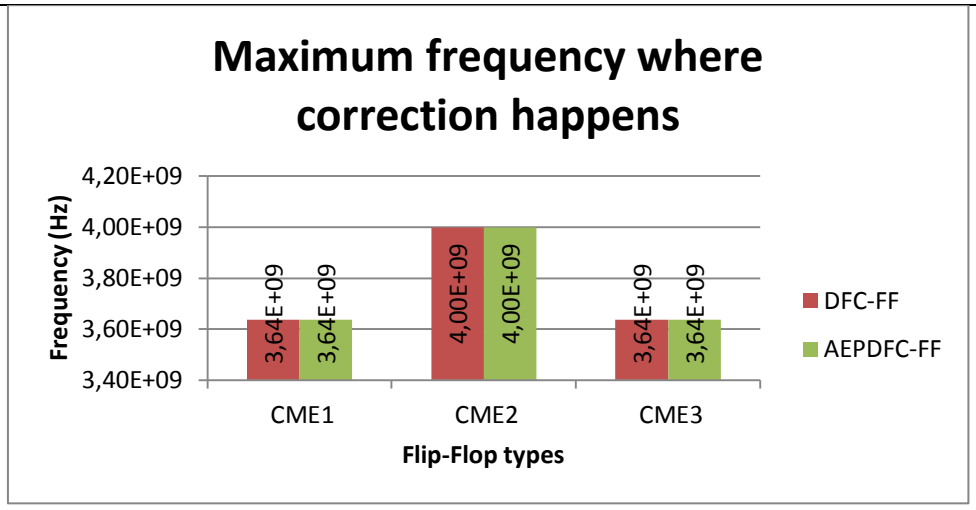
With AgingCalc software the 3 most critical memory elements were identified, in the 2-stage, 4-bit pipeline multiplier. AgingCalc is a software tool that analyzes and predicts the circuit's aging induced by NBTI. AgingCalc evaluates how the threshold voltages of individual transistors vary with time, calculates circuit paths delays, identifies critical memory elements (flip-flops), and generates SPICE netlists for different aging moments in time. The development of this tool started in 2010 at

University of Algarve as part of Jackson Pachito's master thesis, with Prof. Jorge Semião as tutor and it was released in 2011 [3].

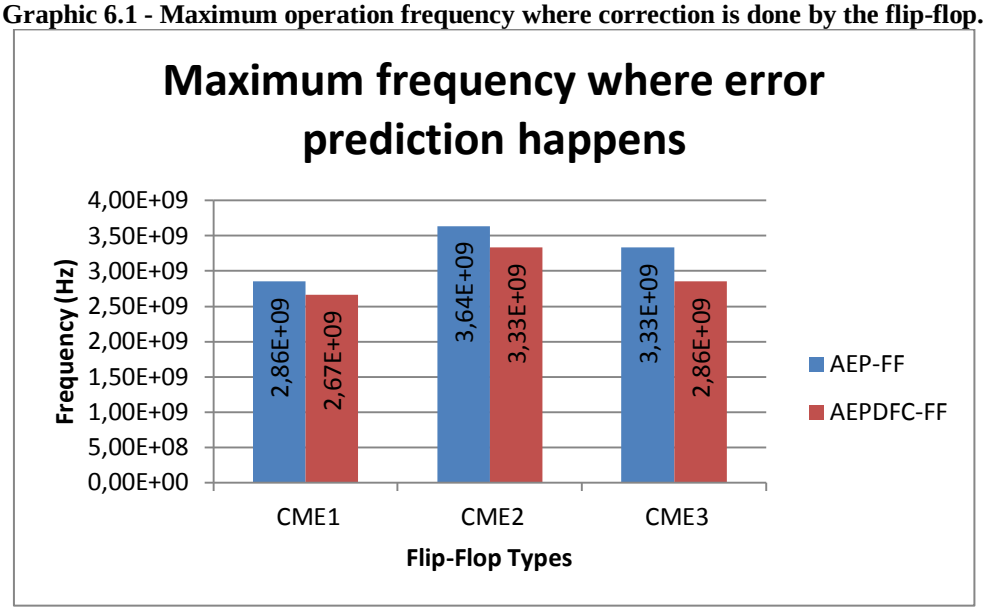
The three most critical memory elements identified are the flip-flops named: "clk_r_REG1_S2" (CME1), "clk_r_REG12_S2" (CME2) and "clk_r_REG11_S2" (CME3). Monitoring the data input signal in each one allows to uncover abnormal delays on up to 27 critical paths (1 CP for CME1, 15 CP for CME2 and 11 CP for CME3), on 636 circuit paths which represent 4.2% path coverage [19].

The three CME were replaced by three AEP-FF, by three DFC_FF and by three AEPDFC-FF. It was decided to use the delay elements developed for the AEP-FF described in [19]. In [19] were developed three delay elements: (i) one with low delay (DE_L); (ii) one with medium delay (DE_M), (iii) and one with high delay (DE_H). The delay elements used in the AEP-FF were the DE_M. The delay elements used in the CES of the DFC_FF were the DE_H. The delay elements used in the AEPDFC-FF were the DE_M and for the AEPDFC-FF CES were used the DE_H.

HSpice simulations were performed with a sweep of the clock signal period between 250ps and 400ps with a 25ps steps. For CME1 the AEP-FF, prediction happens until 2.86GHz; for the AEPDFC-FF prediction happens until 2.67GHz and correction happens until 3.64GHz and for the DFC-FF correction happens also until 3.64GHz. For CME2, the AEP-FF, prediction happens until 3.64GHz, for the AEPDFC-FF prediction happens until 3.33GHz and correction happens until 4.00GHz and for the DFC-FF correction happens also until 4.00GHz. For CME3, the AEP-FF, prediction happens until 3.08GHz, for the AEPDFC-FF prediction happens until 2.86GHz and correction happens until 3.64GHz and for the DFC-FF correction happens also until 3.64GHz. (See Graphic 6.1 and Graphic 6.2)



Graphic 6.1 - Maximum operation frequency where correction is done by the flip-flop.



Graphic 6.2 - Maximum frequency where error prediction is done by the flip-flop.

Despite the new AEPDFC-FF loses for the AEP-FF in error prediction (Graphic 6.2), the possibility of continuing correcting the signal for higher frequencies is a great advantage. The reason why the AEPDFC-FF stops to predict errors sooner than the AEP-FF is explained by the internal architecture of the both Flip-Flops. As it can be observed in Figure 4.5 and Figure 5.3, the aging sensor receives the signal in node Z. In the AEPDFC-FF the clock signal has a delay on the master latch (CTRL signal) and this will delay the arriving of the D signal in node Z. Because the aging sensor is activated with the clock (CLK) signal, this delay will reduce the prediction window.

Despite this, the AEPDFC-FF and the DFC-FF show to be a better solution for faster circuits, because they start correcting the late signals (see Figure 6.2). The AEP-FF does not have the correction late signals characteristic.

Figure 6.2 show that the DFC-FF (third graphic) and the AEPDFC-FF (second graphic) corrects the signal that arrives after the clock signal, while the AEP-FF (forth graphic) does not. Also, neither the AEP-FF neither the AEPDFC-FF (first graphic) perform an error prediction at 3,57GHz.

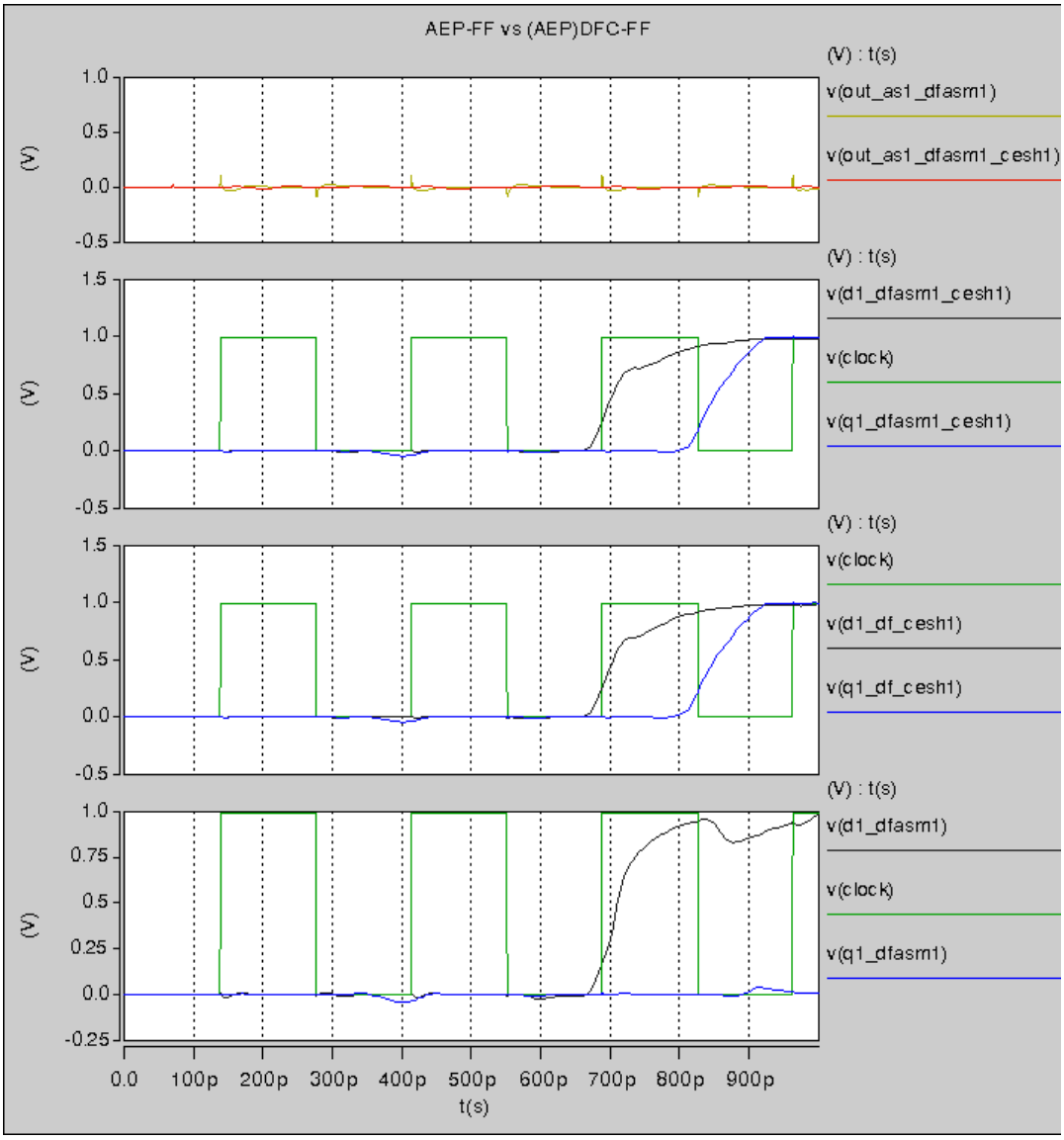


Figure 6.2 – AEP-FF failures vs (AEP)DFC-FF delay fault correction capability.

6.2 V/F CONTROLLER MECHANISM

6.2.1 SIMULATIONS CONDITIONS

The power or frequency optimization controller (V/F Controller) (Figure 4.8) was developed in VHDL, a hardware description language (VHDL – *VHSIC [Very High Speed Integrated Circuits] Hardware Description Language*). This description hardware language was designed to develop integrated Circuits. In fact, VHDL can be used to describe several system types, a computer network, an integrated circuit or a simple logic port.

The working logic of the power or frequency optimization controller is described in chapter 4.3. Also Figure 4.8 shows the state machine logic of the controller.

The VHDL code of the power or frequency optimization controller is listed in appendix A1.

6.2.2 SPICE NETLIST

After the simulations in VHDL confirm that the V/F Controller was working as expected, the VHDL code was converted to Verilog code using INESC-ID software and further using AgingCalc software developed in the University of Algarve the Verilog code was converted in Spice Netlist.

The result is presented in appendix A2.

Tests were performed in HSpice to verify the correct operation of the V/F Controller, presented as a sub circuit in appendix A2.

After these previous tests, the full circuit was assembled to demonstrate the on-line operation of the V/F Controller. The simulation conditions are presented in next subsection.

6.2.3 SIMULATIONS CONDITIONS

The pipeline multiplier was also used as the CUT (Circuit Under Test) to demonstrate the on-line operation of the power or frequency optimization controller, for long-term operation, using global and local performance sensors. For these simulations, the local sensors used were the AEP-FF developed in [19], and because the technology used was 65 nm, the global sensors used was just with the NORs chain where the NBTI effect is the dominant effect.

	CONDITIONS
TECHNOLOGY MODEL (ηm)	PTM 65
FREQUENCY RANGE (GHz)	2.2 to 3
VDD (V)	1.1 (with a step of minus 10%)
TEMPERATURE (ºC)	27

Table 6.2 – Simulation Conditions.

6.2.4 ONLINE OPERATION OF THE VF OPTIMIZATION CONTROLLER

Simulations were performed using Table 6.2 simulation conditions. The circuit under test is the pipeline multiplier and simulations were run for 72ns.

In the beginning of the simulation a reset signal was activated in order to set the VFcode and GSOcode to the default values. These values are acquired through factory tests with the global sensor and local sensors calibration. In this simulation, the VFcode and GSOcode were set to “100” and “011” respectively. The propose is to observe the improvement in the frequency operation. The entry of the DCO is the VFcode, and the “100” code corresponds to a frequency of approximately 2.2GHz Figure 6.4 - VFcode + GSOcode + Code Read from Global Sensor + Local Errors during VFController online operation.(Figure 6.3 and Figure 6.4).

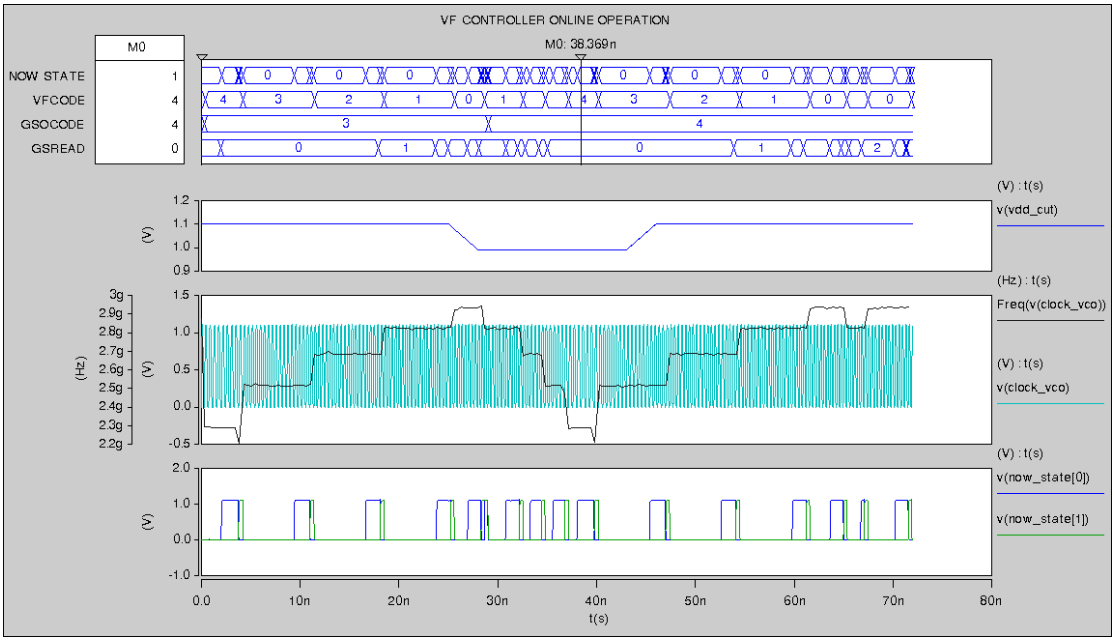


Figure 6.3 - Online VF Controller behavior due to VDD variations

Because there are no local errors and the reading of the global sensor(GSOREAD) is lower than the GSOcode, the VF Controller maximizes the performance by raising the frequency (VFcode is lower). This is a cyclic action, so the frequency is raised while there are no local errors and until reading of the global sensor (GSOREAD) is equal to the GSOcode.

In the simulation, and to validate the correct behavior of the VF Controller, the CUT's V_{DD} was changed to a value 90% of its nominal value, to emulate an aged circuit or a transient delay-fault (Figure 6.3). For this reason, local errors start to be signalized (the error didn't actually occurred, but the AEP-FF predicts the occurrence of the error), and the VF Controller takes an action to reduce frequency (raising VFcode) (Figure 6.3 and Figure 6.4)

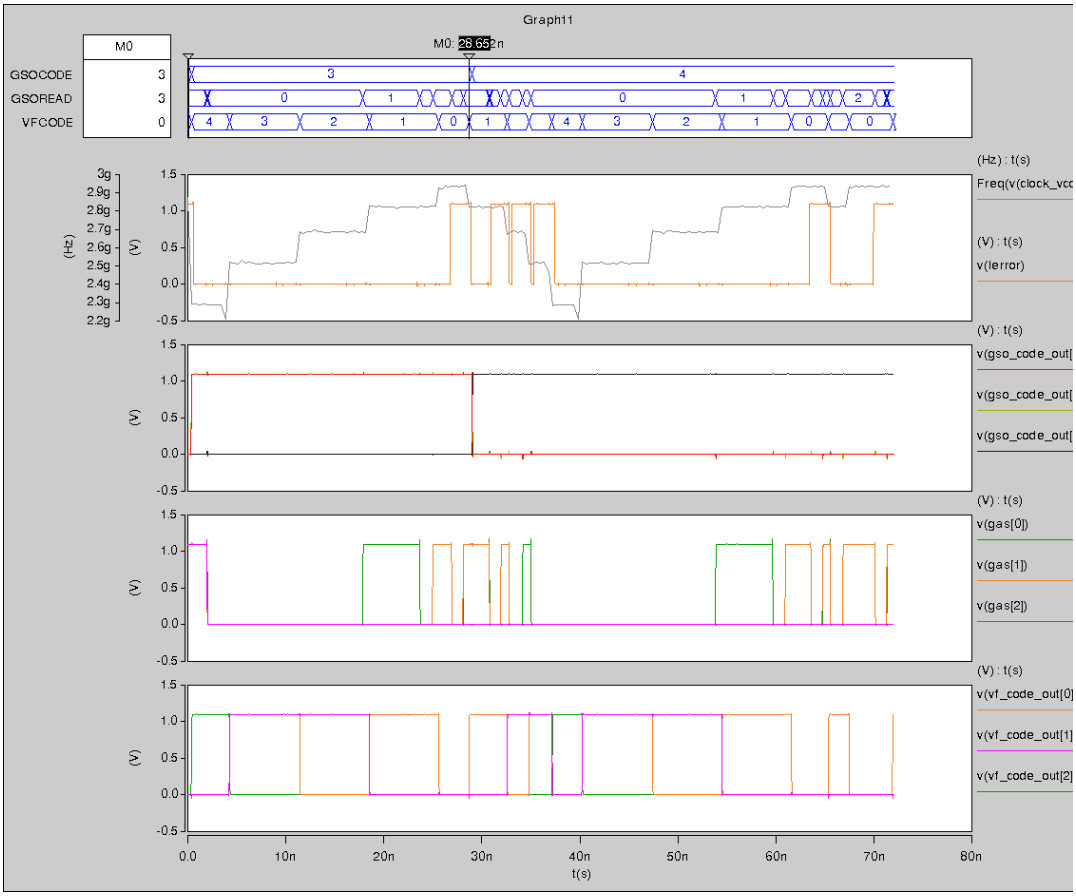


Figure 6.4 - VFcode + GSOcode + Code Read from Global Sensor + Local Errors during VFController online operation.

Again, this is a cyclic action, so the frequency is reduced until there are no local errors and the reading of the global sensor (GSOREAD) is equal to the GSOcode.

Because in this simulation the V_{DD} was raised again to the initial value, the VF Controller reacts to maximize CUT performance by raising again the frequency (decreasing VFcode).

As mentioned in section 3.4, it is expected that each path will suffer delay time variations induced by PVTA variations during circuit operation. The VF controller mechanism is a novel and innovating solution to deal with the problems caused by any of the PVTA variations. The VF controller mechanism allows the circuit to work at maximum performance without circuit's operation malfunction and failure (Delay Fault).

It should be reminded errors do not occur, as delay-fault prediction is used. This way, the controller itself can have a CP larger than the circuit's CP, and the solution is simply to use local sensors in the controller, as if it was part of the CUT. As no errors occur, the controlling process deals with the CP in the controller itself as if it were in the CUT. This is a major novelty regarding available DVFS methodologies.

7. CONCLUSIONS AND FUTURE WORK

7.1 CONCLUSIONS

This thesis focused on the development of a control mechanism of power or frequency optimization, for long-term operation, using global and local performance sensors. Additionally a delay fault correction sensor for synchronous digital circuits was developed as an improvement of the AEP-FF.

The downsizing of new technologies leads to exponential raise on variability and IC sensitivity to disturbances, externals and internals. The scale reduction of devices to nanometric dimensions makes process variability a high contributor to circuit's reliability problems. Also, circuit aging is responsible for longer propagation times in the internal combinational paths, and degrades circuit's performance. Among all variability effects, PVTA variations are main contributors for the performance loss in nanometer technologies. Process variations are static deviations that can put a circuit in a slow design-corner. High temperature operation, namely due to power dissipation, is responsible for boosting aging degradations, and lower power-supply voltage values, namely due to switching activity, is responsible for slow-down circuit's performance.

In this work, the main goal was to develop a new DVS/DFS control mechanism, working with predictive global and local performance sensors, aiming at the optimization of the circuit during a long-term operation (due to their aging aware adaptive sensibility). Therefore, the dynamic voltage or frequency optimization methodology is aging-aware and controls performance during circuit's lifetime with two possible needs: reduce power consumption (controlling the circuit's V_{DD}) or increase frequency operation (controlling the circuit's clock). Global and local performance sensors are used to monitor on-line circuit's timing response, and they complement each other on creating optimized and fail-safe circuits. The identification

of signal paths in the CUT that age faster is relevant, to allow reduced sensor insertion.

The methodology was demonstrated through extensive SPICE simulations, at transistor level, showing that:

1) It is possible to control and implement a dynamic tuning procedure, to control the voltage or frequency self-tuning during circuit's lifetime, based on predictive fault-detection;
2) The developed controller can effectively tune the maximum clock frequency, to improve performance, or to reduce VDD voltage, to improve power dissipation, maintaining a fault-free operation;
3) The control mechanism for the tuning procedure, as it is based on predictive-fault detection sensors, i.e., the fault presumably never occurs, can monitor their own critical path, and thus the fault-detection circuit is treated as part of the CUT, itself, just by implementing local sensors on the control circuitry, if necessary;

An additional goal in this work was to deal with variability and develop a DFC-FF, in order to maintain performance under high variability. If a delay variation in a path imposes its delay to surpass the clock period, the error will be corrected by the DFC-FF and the lost reliability is recovered. Two architectures were presented for the new flip-flops: (1) A Delay-Fault Correction type, to correct the CP that exceeds the clock period's standard time; (2) a Delay-Fault Detection and Correction FF with Predictive fault detection, to allow simultaneously the detection and correction of delay-faults locally, in the critical memory elements where they occur. The former FF is a big boost in circuit's dependability and reliability. However, the area overhead and FF complexity are the main drawbacks. Nevertheless, the AEPDFC-FF can be used in the previously DVS/DFS self-tuning methodology due to its error-prediction capability. The error correction is an additional ultimate protection, if increased PVTA variations occur.

For high performance circuits, this error correction circuitry revealed to have limitations, which imposed the definition a Fast-DFC architecture. This fast error correction circuitry has reduced circuit complexity and, simultaneously fewer transistors count. However, it has the drawback of putting the master-latch in

transparent mode, during the correction window, which takes place in the active part of the clock. This procedure, may allow the transmission of glitches and crosstalk from one cycle to the next one, and therefore a more thorough analysis is needed in this cell in future work.

7.2 FUTURE WORK

Every research work is an unfinished task and many future perspectives are now open for this work.

First, a test-chip is mandatory, to analyze real silicon data and to validate on silicon the proposed methodologies and circuits.

Regarding the DVS/DFS methodology, the controller operation can be optimized, to reduce the number of clock cycles to perform the complete tuning procedure. Also, the global sensor may also be optimized, to simplify the testing vectors and, consequently, reduce the number of clock periods necessary to perform a global sensing procedure. Additionally, the global sensors' cell may be simplified and replaced by a simple latch, reducing area overhead.

Regarding the DFC-FFs, integration in the generic CMOS library used by AgingCalc will allow an aging analysis and aging prediction. As already mentioned, a more thorough analysis on the Fast-DFC-FF is needed, to understand the implications of crosstalk and glitches on its reliability.

Finally, integrating the AEPDFC-FF with the DVS/DFS methodology will allow to create a robust solution for on-line optimization.

REFERENCES

[1] Oliveira, R. S.; Semião, J.; Teixeira, I.C.; Santos, M.B.; Teixeira, J.P., "On-line BIST for performance failure prediction under aging effects in automotive safety-critical applications," Test Workshop (LATW), 2011 12th Latin American , vol., no., pp.1,6, 27-30 March 2011

[2] X. Fu, T. Li, J. Fortes, "NBTI Tolerant Microarchitecture Design in the Presence of Process Variation", Proc. 41st Annual IEEE/ACM International Symposium on Microarchitecture, IEEE Computer Society, Washington, DC, USA, pp. 399-410, DOI: 10.1109/MICRO.2008.4771808, 8-12 Nov, 2008.

[3] J. Pachito, "Metodologia para prever o envelhecimento de circuitos digitais", Master of Science Thesis, Universidade do Algarve – Instituto Superior de Engenharia, 2011.

[4] Vazquez, J.C.; Champac, V.; Ziesemer, A.M.; Reis, R.; Semiao, J.; Teixeira, I.C.; Santos, M.B.; Teixeira, J.P., "Predictive error detection by on-line aging monitoring," On-Line Testing Symposium (IOLTS), 2010 IEEE 16th International , vol., no., pp.9,14, 5-7 July 2010

[5] C. Martins, J. Semião, J. C. Vazquez, V. Champac, M. Santos, I. C. Teixeira, J. P. Teixeira, "Adaptive Error Prediction Flip-Flop for Performance Failure Prediction with Aging Sensors", Proc. of IEEE VLSI Test Symposium (VTS), April 2011.

[6] H. Abrishami, S. Hatami, B. Amelifard, M. Pedram, "Multi-Corner, Energy-Delay Optimized, NBTI-Aware Flip-Flop Characterization and Design", Proc. 18th ACM Great Lakes Symposium on VLSI, ACM New York, NY, USA, pp. 652-659, DOI: 10.1109/ISQED.2010.5450509, 22-24 March, 2010.

[7] F. Marc, B. Mongellaz, C. Bestory, H. Levi, Y. Danto, "Improvement of Aging Simulation of Electronic Circuits Using Behavioral Modeling", Tran. Device and Materials Reliability, IEEE Electron Devices Society and IEEE Reliability

Society, Vol. 6 Issue: 2, pp. 228-234, DOI: 10.1109/TDMR.2006.879117, June 2006.

[8] J. Semião, "Power-Supply and Temperature Based Methodologies to Improve Tolerance and Detection of Delay Faults in Synchronous Digital Circuits", Ph.D. Thesis, IST, TUL, July, 2010.

[9] V. Pallipadi, "Enhanced Intel SpeedStep® Technology and Demand-Based Switching on Linux", Intel®, 05/10/2010, http://software.intel.com/en-us/articles/enhanced-intel-speedstepr-technology-and-demand-based-switching-on-linux/

[10] "AMD PowerNow!™ Technology Dynamically Manages Power and Performance - Informational White Paper", AMD, Publication: 24404 Rev: A Amendment/0, November 2000. - http://www.amd-k6.com/wp-content/uploads/2012/07/24404a.pdf

[11] Shidhartha Das, et al, "A Self-Tuning DVS Processor Using Delay-Error Detection and Correction", Journal of Solid State Circuits, Vol. 41, NO. 4, April, 2006.

[12] D. Blaauw, S. Kalaiselvan, K. Lai, Wei-Hsiang Ma, S. Pant, C. Tokunaga, S. Das, D. Bull, "Razor II: In Situ Error Detection and Correction for PVT and SER Tolerance," Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International , vol., no., pp.400-622, 3-7 Feb. 2008.

[13] J. Semião et al., "Time Management for Low-Power Design of Digital Systems", ASP Journal of Low Power Electronics (JOLPE), Vol. 4, N° 3, pp. 410-419, Dec., 2008.

[14] J. C. Vazquez et al., "Built-In Aging Monitoring for Safety-Critical Applications", Proc. IEEE Int. On-Line Test Symp. (IOLTS), pp. 9-14, 2009.

[15] J. C. Vazquez et al., "Low-sensitivity to Process Variations Aging Sensor for Automotive Safety-Critical Applications", Proc. IEEE VLSI Test Symposium (VTS), pp. 238-243, 2010.

[16] C. V. Martins, J. Pachito, J. Semião, I. C. Teixeira, J. P. Teixeira, "Adaptive Error-Prediction Aging Sensor for On-Line Monitoring of Performance Errors", Proceedings of the XXVI Conference on Design of Circuits and Integrated Systems - DCIS'11, Albufeira, Portugal, November, 2011.

[17] S. V. Kumar, C. H. Kim, S. S. Sapatnekar, Fellow, "Adaptive Techniques for Overcoming Performance Degradation Due to Aging in CMOS Circuits", IEEE

Transactions on Very Large Scale Integration (VLSI) Systems, vol. 19, no. 4, pp. 603-614, April 2011

[18] E. Takeda, C. Y. Yang, and A. Miura-Hamada, “Hot-Carrier Effects in MOS Devices”, Academic Press, 1995.

[19] C. Martins, "Adaptive Error-Prediction Aging Sensor for Synchronous Digital Circuits ", Msc. Thesis, ISE, UALG, October, 2012.

[20] http://www.siliconfareast.com/emig.htm, July 2013.

[21] J. Vazquez, V. Champac, A. Ziesemer, R. Reis, I. Teixeira, M. Santos, and J. Teixeira, “Low-sensitivity to process variations aging sensor for automotive safety critical applications”, April 2010.

[22] N. Weste, D. Harris, “CMOS VLSO Design – a Circuits and Systems Perpective”, Pearson Education, Inc., publishing as Addison-Wesley, ISBN 13: 978-0-321-54774-3

[23] M. Agarwal, B. Paul, M. Zhang, and S. Mitra, “Circuit Failure Prediction and Its Application to Transistor Aging”, Proc. VLSI Test Symp. (VTS), pp. 277-286, May 2007.

[24] G. Jovanovi´c, M. Stoj˘cev, Z. Stamenkovic “A CMOS Voltage Controlled Ring Oscillator with Improved Frequency Stability”, Scientific Publications of the state university of Nozi Pazar, SER. A: Appl. Math Inform. and Mech., May 2010, Vol. 2, 1, pp 1-9.

[25] J. Semiao, J. Pachito, C. Martins, B. Jacinto, J. Vazquez, V. Champac, M. Santos, I. Teixeira, J. Teixeira, "Aging-aware Power or Frequency Tuning with Predictive Fault Detection", IEEE Design & Test of Computers, Volume 29 , Issue 5, September/October 2012, DOI: 10.1109/MDT.2012.2206009.

[26] J. Semião, M. Rodriguez-Irago, L. Piccoli, F. Vargas, M. B. Santos, I. C. Teixeira, J. J. Rodríguez-Andina, J. P. Teixeira, “Signal Integrity Enhancement in Digital Circuits”, IEEE Design and Test of Computers, September-October 2008, vol. 25, no. 5, pp. 452-461

[27] D. Kim, J. Kim, M. Kim, J. Moulic, H. Song, "System and Method for Monitoring Reliability of a Digital System", IBM Corp., US Patent 7495519, Feb. 24, 2009.

[28] Predictive Technology Model (PTM), http://www.eas.asu.edu/~ptm/.

[29] Hans Jacobson, Pradip Bose, Zhigang Hu, Alper Buyuktosunoglu, Victor Zyuban, Rick Eickemeyer, Lee Eisen, John Griswell, Doug Logan, Balaram

Sinharoy, Joel Tendler, "Stretching the Limits of Clock-Gating Efficiency in Server-Class Processors", 11th International Symposium on High-Performance Computer Architecture (HPCA'05), San Francisco, California, February, 2005, ISBN: 0-7695-2275-0.

[30] Kaijian Shi and David Howard, "Sleep Transistor Design and Implementation - Simple Concepts Yet Challenges To Be Optimum", Proc.. IEEE VLSI-DAT, April, 2006.

[31] J. Semiao, J. Pachito, C. Martins, M. Santos, I. Teixeira, P. Teixeira, "The Influence of Clock-Gating On NBTI-Induced Delay Degradation", 18th IEEE International On-Line Testing Symposium – IOLTS'12, Sitges, Spain, June 27-29, 2012.

[32] K. Kuhn, C. Kenyon, A. Kornfeld, M. Liu, A. Maheshwari, W. Shih, S. Sivakumar, G. Taylor, P. VanDerVoorn, K. Zawadzki, "Managing Process Variation in Intel's 45nm CMOS Technology," Intel® Technology Journal, Volume 12, Issue 02, Published in 17/06/2008. http://www.intel.com/technology/itj/2008/v12i2/3-managing/1-abstract.htm.

[33] M. Keating, D. Flynn, R. Aitken, A. Gibbons, K. Shi, "Low Power Methodology Manual for System on Chip Design", Springer Publications, New York, 2007, ISBN-13: 9780387718187, ISBN: 0387718184

[34] "Design for Variability: Managing Design, Process, and Manufacturing Variations in Physical Design", Mentor Graphics Corporation, Place & Route Whitepaper, 2008. - http://www.mentor.com/resources/techpubs/upload/mentorpaper_43548.pdf.

[35] Masaoud Houshmand Kaffashian, Reza Lotfi, Khalil Mafinezhad, Hamid Mahmoodi, "Impacts of NBTI/PBTI on performance of domino logic circuits with high-k metal-gate devices in nanoscale CMOS", Microelectronics Reliability, Volume 52, Issue 8, August 2012, Pages 1655-1659, ISSN 0026-2714.

[36] S. Pae, J Maiz, C. Prasad, B Woolery, "Effect of BTI Degradation on Transistor Variability in Advanced Semiconductor Technologies", Device and Materials Reliability, IEEE Transactions on , vol.8, no.3, pp.519,525, Sept. 2008

[37] C. V. Martins, J. Semião, I. C. Teixeira, J. P. Teixeira, "Flip-Flop com Sensor para Previsão de Erros de Performance, para Aplicação como Sensor Local de Performance ou de Envelhecimento em Circuitos Integrados Digitais

Síncronos", Patente Portuguesa - PT 105636, Universidade do Algarve, 20 de Abril de 2011.

[38] J. Semião, I. C. Teixeira, J. P. Teixeira, A. Romão, "Optimização ao Longo da Vida da Dissipação de Potência ou da Performance em Circuitos Integrados Digitais Síncronos", Patente Portuguesa - PT 106513, Universidade do Algarve, 27 de Agosto de 2013.

[39] K. Kang, S. Gangwal, S. Phil Park, and K. Roy, "NBTI Induced Performance Degradation in Logic and Memory Circuits: How Effectively Can We Approach a Reliability Solution?", Proc. Asia / South Pacific Design Autom. Conf. (ASP-DAC), pp. 726-731, 2008.

[40] K Soumyanath, S. Borkar, Z. Chunyan, and B.Bloechel, "Accurate on-chip interconnect evaluation: a time-domain technique." IEEE Journal of Solid-State Circuits, Volume 34 Issue, May 1999, pp. 623–631.

[41] V. Mehrotra, "Modeling the effects of manufacturing variation on high-speed microprocessor interconnect performance." IEEE International Electron Devices Meeting, IEDM Technical Digest, December 1998, pp.767–770.

APPENDIX

A1. VHDL CODE

```
------------------------------------------------------------
LIBRARY ieee;
USE ieee.std_logic_1164.all;

ENTITY controlador IS
  PORT (
    Resetn, LError, Clock, Timer, Global_Sensing_End : IN
STD_LOGIC;
    GSensor_read : IN STD_LOGIC_VECTOR(2 downto 0);
add_VF_code, sub_VF_code, add_GSO_code, VF_code_reset,
GSO_code_reset : in STD_LOGIC_VECTOR(2 downto 0); --
VF_code_reset="011"
-- GSO_code_reset="010"
    GSensor_read_equality, GSensor_read_bigger : in
STD_LOGIC;
    VF_code_out : out STD_LOGIC_VECTOR(2 downto 0);
     GSO_code_out : out STD_LOGIC_VECTOR(2 downto 0);
     Global_sense : OUT STD_LOGIC;    Reset_LSensorN : OUT
STD_LOGIC
    );
  END controlador;

ARCHITECTURE Behavior OF controlador IS

  TYPE State_type IS (Local_Sensing, Global_Sensing,
ChangeVF, UpdateGSO);
  SIGNAL now_state, next_state : State_type;
  SIGNAL VF_code, next_VF_code: STD_LOGIC_VECTOR(2 downto
0);
  SIGNAL GSO_code, next_GSO_code : STD_LOGIC_VECTOR(2 downto
0);
  SIGNAL Reset_LSensorN_next : STD_LOGIC;

BEGIN

  Maquina_de_Estados: PROCESS ( Resetn, GSO_code_reset,
VF_code_reset, GSO_code, VF_code, Timer, LError,
Global_Sensing_End, GSensor_read_bigger, sub_VF_code,
add_VF_code, GSensor_read_equality)
  BEGIN
```

```
    IF Resetn = '0' THEN
        next_GSO_code <= GSO_code_reset;
        next_VF_code <= VF_code_reset;
        next_state <= Local_Sensing;
        Global_sense <= '0';
        Reset_LSensorN_next <= '0';
    ELSIF now_state = Local_Sensing THEN
        next_GSO_code <= GSO_code;
        next_VF_code <= VF_code;
        next_state <= Local_Sensing;
        Global_sense <= '0';
        Reset_LSensorN_next <= '1';
        IF Timer = '1' OR LError ='1' THEN
          next_state <= Global_Sensing;
          Global_sense <= '1';
        END IF;

    ELSIF now_state = Global_Sensing THEN
        next_GSO_code <= GSO_code;
           next_VF_code <= VF_code;
           next_state <= Global_Sensing;
           Global_sense <= '1';
        Reset_LSensorN_next <= '1';
        IF (Global_Sensing_End = '1') THEN
          IF (GSensor_read_equality = '1') THEN
         IF (LError ='0') THEN)
              next_state <= Local_Sensing;
            ELSE
              next_state <= ChangeVF;
         END IF;
          ELSE
            next_state <= ChangeVF;
          END IF;
        END IF;

    ELSIF now_state = ChangeVF THEN
        next_GSO_code <= GSO_code;
        next_VF_code <= VF_code;
        next_state <= Local_Sensing;
        Global_sense <= '0';
        Reset_LSensorN_next <= '1';
        IF (LError ='1') THEN
       if (GSensor_read_equality = '1') then
            next_VF_code <= add_VF_code;
            next_state <= UpdateGSO;
       else
            next_VF_code <= add_VF_code;
       end if;
          Reset_LSensorN_next <= '0';
     ELSE
          IF (GSensor_read_bigger = '0') THEN
```

```
                next_VF_code <= sub_VF_code;
            ELSE
                next_VF_code <= add_VF_code;
            END IF;
          END IF;

      ELSE
          next_GSO_code <= add_GSO_code;
             next_VF_code <= VF_code;
             next_state <= Local_Sensing;
             Global_sense <= '0';
          Reset_LSensorN_next <= '0';
      END IF;
    END PROCESS;

    VF_code_out <= VF_code;
    GSO_code_out <= GSO_code;

    -- Processo de Sincronismo / Reset
    Sic_reg: PROCESS ( Clock, Resetn)
    BEGIN
      IF (Clock'EVENT AND Clock = '1') THEN
        now_state <= next_state;
        VF_code <= next_VF_code;
        GSO_code <= next_GSO_code;
        Reset_LSensorN <= Reset_LSensorN_next;
      END IF;
    END PROCESS;

END Behavior;
```

A2. SPICE NETLIST

```
.Subckt ControllerVF
 +GSensor_read[2] GSensor_read[1] GSensor_read[0]
 +add_VF_code[2] add_VF_code[1] add_VF_code[0]
 +sub_VF_code[2] sub_VF_code[1] sub_VF_code[0]
 +add_GSO_code[2] add_GSO_code[1] add_GSO_code[0]
 +VF_code_reset[2] VF_code_reset[1] VF_code_reset[0]
 +GSO_code_reset[2] GSO_code_reset[1] GSO_code_reset[0]
 +Resetn
 +LError
 +Clock01 Clock02 Clock03 Clock04 Clock05 Clock06 Clock07 Clock08 Clock09
 +Timer
 +Global_Sensing_End
 +GSensor_read_equality
 +GSensor_read_bigger
 +VF_code_out[2] VF_code_out[1] VF_code_out[0]
 +GSO_code_out[2] GSO_code_out[1] GSO_code_out[0]
 +Global_sense
 +Reset_LSensorN
 +Vss Vdd


X_now_state_reg[0] next_state[0] Clock01 now_state[0] n41 Vss Vdd DF1
X_now_state_reg[1] next_state[1] Clock02 now_state[1]
node__now_state_reg[1]_QN Vss Vdd DF1
X_VF_code_reg[2] next_VF_code[2] N80 Clock03 VF_code_out[2] n47 Vss Vdd
DFE1
X_VF_code_reg[1] next_VF_code[1] N80 Clock04 VF_code_out[1] n46 Vss Vdd
DFE1
X_VF_code_reg[0] next_VF_code[0] N80 Clock05 VF_code_out[0] n45 Vss Vdd
DFE1
X_GSO_code_reg[2] next_GSO_code[2] N80 Clock06 GSO_code_out[2] n44 Vss Vdd
DFE1
X_GSO_code_reg[1] next_GSO_code[1] N80 Clock07 GSO_code_out[1] n43 Vss Vdd
DFE1
X_GSO_code_reg[0] next_GSO_code[0] N80 Clock08 GSO_code_out[0] n42 Vss Vdd
DFE1
XReset_LSensorN_reg Reset_LSensorN_next Clock09 Reset_LSensorN
node_Reset_LSensorN_reg_QN Vss Vdd DF1
XU44 n49 n50 n51 n52 next_state[1] Vss Vdd OAI310
XU45 LError n53 n50 Vss Vdd NOR20
XU46 GSensor_read_equality n53 Vss Vdd CLKIN0
XU47 Global_Sensing_End n49 Vss Vdd CLKIN0
XU48 Global_Sensing_End n51 n54 n52 next_state[0] Vss Vdd OAI2110
XU49 n55 GSensor_read_equality LError n52 Vss Vdd NAND30
XU50 n56 n47 n57 n58 next_VF_code[2] Vss Vdd OAI2110
```

```
XU51 sub_VF_code[2] n59 add_VF_code[2] n60 n58 Vss Vdd AOI220
XU52 VF_code_reset[2] n61 n57 Vss Vdd NAND20
XU53 n56 n46 n62 n63 next_VF_code[1] Vss Vdd OAI2110
XU54 sub_VF_code[1] n59 add_VF_code[1] n60 n63 Vss Vdd AOI220
XU55 VF_code_reset[1] n61 n62 Vss Vdd NAND20
XU56 n56 n45 n64 n65 next_VF_code[0] Vss Vdd OAI2110
XU57 sub_VF_code[0] n59 add_VF_code[0] n60 n65 Vss Vdd AOI220
XU58 n66 n60 Vss Vdd CLKIN0
XU59 GSensor_read_bigger LError n55 n66 Vss Vdd OAI210
XU60 n67 n55 Vss Vdd CLKIN0
XU61 GSensor_read_bigger LError n67 n59 Vss Vdd NOR30
XU62 VF_code_reset[0] n61 n64 Vss Vdd NAND20
XU63 n68 n56 Vss Vdd CLKIN0
XU64 n67 n44 n69 n70 next_GSO_code[2] Vss Vdd OAI2110
XU65 GSO_code_reset[2] n61 n70 Vss Vdd NAND20
XU66 add_GSO_code[2] n68 n69 Vss Vdd NAND20
XU67 n67 n43 n71 n72 next_GSO_code[1] Vss Vdd OAI2110
XU68 GSO_code_reset[1] n61 n72 Vss Vdd NAND20
XU69 add_GSO_code[1] n68 n71 Vss Vdd NAND20
XU70 n67 n42 n73 n74 next_GSO_code[0] Vss Vdd OAI2110
XU71 GSO_code_reset[0] n61 n74 Vss Vdd NAND20
XU72 add_GSO_code[0] n68 n73 Vss Vdd NAND20
XU73 n61 n41 N80 n68 Vss Vdd OAI210
XU74 LError n67 N80 Reset_LSensorN_next Vss Vdd OAI210
XU75 Resetn n41 now_state[1] n67 Vss Vdd NAND30
XU76 n75 N80 Vss Vdd CLKIN0
XU77 n54 n51 Global_sense Vss Vdd NAND20
XU78 now_state[0] n75 n51 Vss Vdd NAND20
XU79 Timer LError n75 n41 n54 Vss Vdd OAI2110
XU80 n61 now_state[1] n75 Vss Vdd NOR20
XU81 Resetn n61 Vss Vdd CLKIN0

.ends
```

A3. PATENT

PATENTE

MICROELECTRÓNICA

CÓDIGO: PT 106513

OPTIMIZAÇÃO AO LONGO DA VIDA DA DISSIPAÇÃO DE POTÊNCIA OU DA PERFORMANCE EM CIRCUITOS INTEGRADOS DIGITAIS SÍNCRONOS

RESUMO

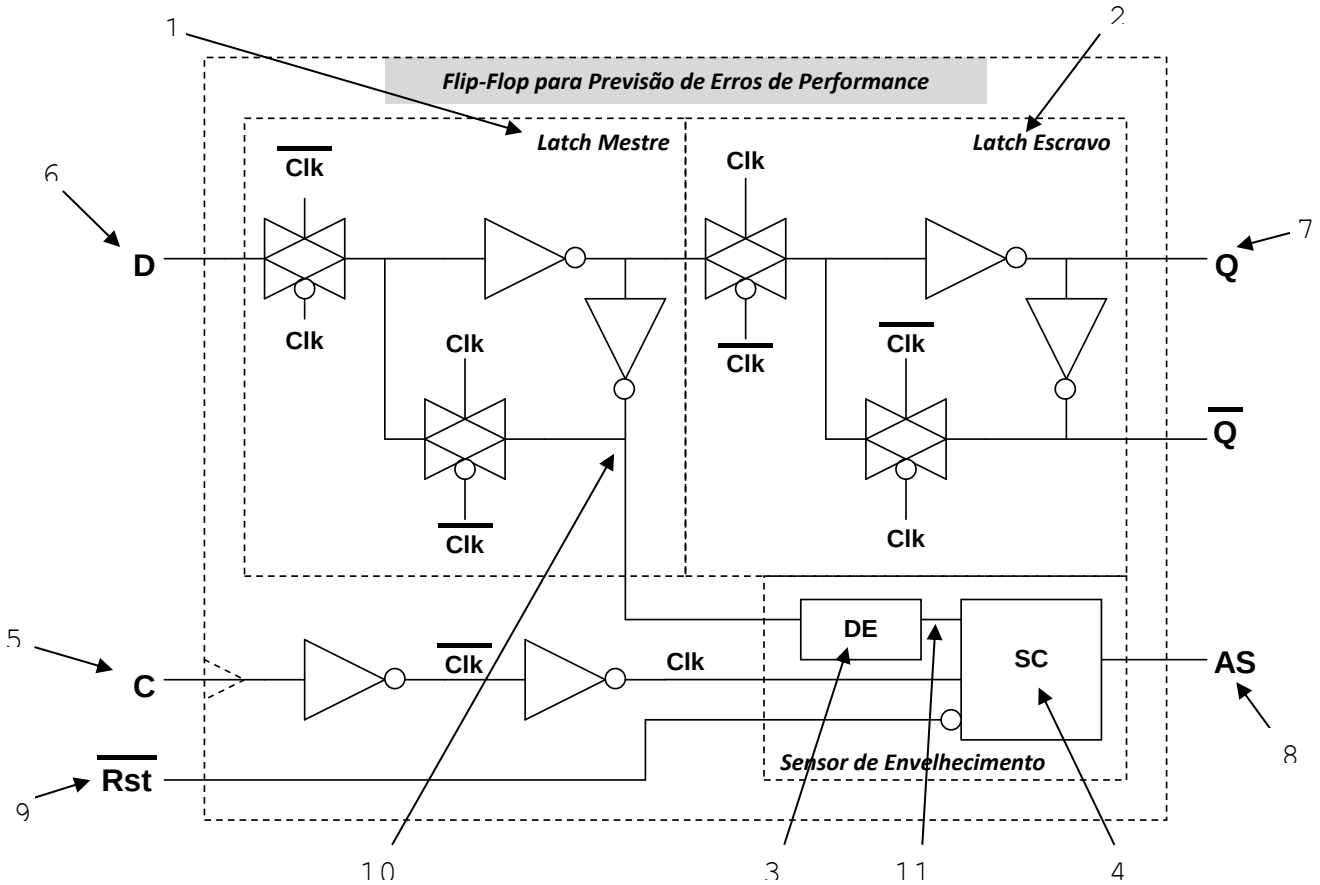
OPTIMIZAÇÃO AO LONGO DA VIDA DA DISSIPAÇÃO DE POTÊNCIA OU DA PERFORMANCE EM CIRCUITOS INTEGRADOS DIGITAIS SÍNCRONOS

A presente invenção refere-se a um sistema para optimizar de forma dinâmica a operação ao longo da vida de circuitos integrados digitais síncronos, permitindo que os circuitos sejam optimizados de acordo com duas necessidades possíveis: (i) restringir a dissipação de potência, reduzindo a tensão de alimentação para o valor mínimo que impede a ocorrência de erros; ou (ii) aumentar o desempenho (performance), aumentando a frequência de operação até ao limite máximo que garante a operação sem erros. O método inclui a arquitectura que permite a optimização possível, constituída por um controlador, um bloco de sensores globais e por vários sensores locais localizados em determinados flip-flops do circuito. Os sensores globais são utilizados para realizar a monitorização periódica (a pedido) dos atrasos no circuito digital. Os sensores locais são utilizados para realizar a monitorização constante dos atrasos nos caminhos mais longos do circuito, sempre que estes são activados, e para calibrar os sensores globais. Todos os sensores são sensíveis a variações PVTA (variações no Processo de fabricação do circuito integrado - P, na tensão de alimentação - V, na temperatura – T e variações provocadas pelo Envelhecimento dos circuitos - A) e podem desencadear a sintonia da tensão de alimentação ou da frequência de operação, durante a vida do produto (circuito integrado digital). A monitorização efectuada pelo bloco de sensores globais fornece uma avaliação grosseira do estado de funcionamento do circuito,

enquanto os sensores locais, quando activados, fornecem uma avaliação fina sobre qual a performance do circuito para evitar erros funcionais. A inserção dos sensores locais é realizada através de uma análise de previsão de envelhecimento, permitindo não só a identificação dos caminhos críticos, mas também os caminhos que envelhecem mais rápido (e que se tornarão críticos com o envelhecimento do circuito). A arquitectura dos sensores propostos é tal que as variações PVTA que ocorrem sobre eles fazem aumentar a sua capacidade de prever o erro, ou seja, os sensores vão-se adaptando ao longo da sua vida útil para aumentarem a sua sensibilidade.

Faro, 27 de Agosto de 2013

FIGURA PARA PUBLICAÇÃO



DESCRIÇÃO

OPTIMIZAÇÃO AO LONGO DA VIDA DA DISSIPAÇÃO DE POTÊNCIA OU DA PERFORMANCE EM CIRCUITOS INTEGRADOS DIGITAIS SÍNCRONOS

1 – Introdução e enquadramento

Os circuitos integrados apresentam durante a sua vida um processo de degradação das suas características iniciais em que vão modificando lentamente o seu comportamento. A esse processo de degradação chamamos de envelhecimento. É um processo lento e cumulativo, em que ocorre um desgaste dos circuitos provocado pelo tempo de operação e condições de funcionamento, com origem em diversos factores, como temperatura, tensão de alimentação, humidade, frequência de operação, radiação exterior presente, etc. [1][2][3]. No fundo, o envelhecimento é provocado por todos os mecanismos que acabam por alterar os parâmetros físicos e eléctricos dos circuitos, diminuindo o seu tempo de vida útil [4].

Para além disso, este envelhecimento traduz-se normalmente no aumento do tempo de propagação dos caminhos, associado à menor condutividade média dos transístores. Enquanto o aumento dos tempos estiver garantido pelo tempo de segurança (definido normalmente por *time slack*, ou seja a margem de tempo que resta num período, depois de subtraído o tempo de propagação do caminho) existente dentro do período de operação dos circuitos sequenciais, não ocorrerá nenhum erro de sincronismo [5]. No entanto, se o tempo de propagação aumentar, potenciado por variações do processo de fabrico (*Process variations*), ou por variações na tensão de alimentação (*power-supply Voltage variations*), ou por

variações na temperatura de operação (*Temperature variations*), ou pelo envelhecimento (*Aging variations*), ou por um somatório de todos os efeitos (*PVTA variations*), ocorre um erro de sincronismo [6]. As células de memória (normalmente os flip-flops) não capturam o valor correcto, mas um valor atrasado, devido ao aumento do tempo de propagação dos caminhos combinatórios que terminam na entrada da célula de memória [7]. Assim, embora as variações PVT possam provocar os mesmos erros de sincronismo que as variações provocadas pelo envelhecimento, como estas últimas são variações cumulativas, fazem com que o tempo de vida útil diminua, porque os circuitos integrados (CI) digitais começam a ter cada vez mais propensão para falhar, por ficarem mais lentos [3][8][9].

Os circuitos integrados realizados em tecnologias nanométricas apresentam uma maior sensibilidade a quaisquer variações paramétricas que possam ocorrer, como por exemplo, variações PVTA. A existência de transístores mais pequenos, em que a condução é feita por um menor número de iões e electrões, torna os circuitos e componentes mais vulneráveis a influências externas. O aumento da variação de qualquer parâmetro existente, em nano tecnologias conduz a maiores incertezas no comportamento eléctrico dos circuitos e a menores níveis de energia necessários para modificar esse comportamento, ficando os circuitos mais susceptíveis a modificações da sua estrutura e do seu comportamento. O envelhecimento é, portanto, mais acelerado em nano tecnologias [3][10]. Em tecnologias anteriores às nano métricas, este envelhecimento também ocorria, mas como os efeitos sobre os circuitos eram menores, eles não se faziam notar durante a vida útil dos circuitos, estando garantidos

pelos generosos tempos de segurança (*time slacks*) normalmente utilizados. Então, dizia-se que os circuitos, praticamente, não envelheciam. Porém, em nano tecnologias, e mais concretamente no caso dos circuitos integrados digitais, os efeitos do envelhecimento fazem-se notar mais cedo durante a vida útil do circuito e as margens de segurança (*time slack*) utilizadas começam a não ser suficientes para garantir uma vida útil do circuito sem ocorrerem erros de sincronismo nas células de memória (os flip-flops e as células de memória síncronas não conseguem capturar o valor correcto, pois ele chega atrasado em relação ao sinal de relógio, que marca o sincronismo), causadas principalmente por variações PVTA [5]. Em suma, a redução do tamanho da tecnologia conduz, normalmente, a circuitos com menores tempos de vida útil.

Na verdade, como referido em [11], se considerarmos aplicações em nano tecnologias onde a segurança é crítica e onde os circuitos têm tempos médios de utilização relativamente elevados, como é o caso da electrónica para a indústria automóvel, onde os camiões têm tempos médios de vida útil de 15 a 20 anos (e onde a electrónica utilizada já é maioritariamente em tecnologias nano métricas [12]), então a degradação cumulativa provocada pelo envelhecimento começa a tornar-se crítica para o correcto desempenho dos circuitos. É imperativo que o envelhecimento seja considerado na fase de projecto, de forma a prever a sua existência e garantir que os circuitos se mantêm a funcionar durante o tempo de vida útil esperado, sem erros, sobretudo nas novas tecnologias nano métricas [13].

Os efeitos mais importantes que provocam o envelhecimento (*Aging*) dos circuitos são *Bias Temperature Instability*

(BTI) [14][15], *Time Dependent Dielectric Breakdown* (TDDB) [16] e o *Hot Carrier Injection* (HCI) [17][18]. No entanto, os efeitos BTI, com as suas duas vertentes NBTI (Negative BTI) [3] e PBTI (Positive BTI) [33][34][35][36][37] que afectam normalmente a tensão limiar de condução dos transístores PMOS e NMOS, respectivamente, têm sido identificados como os efeitos dominantes no processo de degradação a longo prazo [23][24]. Porém, todas as variações já mencionadas, como as variações PVTA, podem degradar e comprometer a performance de um circuito integrado, e podem conduzir à ocorrência de erros e falhas no circuito, especialmente se vários desses efeitos ocorrerem simultaneamente ou se ocorrem degradações cumulativas.

Para lidar com a variabilidade de parâmetros existente nos circuitos de nano tecnologias, os projectistas utilizam abordagens conservativas, com maiores tempos relativos de segurança (*time slacks*, ou margens de erro), o que conduz a circuitos que não estão optimizados. Se consideramos os circuitos com tempos de vida útil maiores, como os já referidos circuitos electrónicos para a indústria automóvel, entre outros, os tempos de segurança utilizados têm de ser muito generosos, para garantir uma operação sem erros. O resultado são circuitos subaproveitados, com uma performance muito aquém do possível, durante os primeiros anos de operação. Por outro lado, como os tempos de propagação das portas num circuito digital são inversamente proporcionais à tensão de alimentação existente (já que maiores tensões de alimentação fazem aumentar a condução nos transístores), é possível também melhorar a gestão de energia nos circuitos que apresentem margens de segurança grandes à tensão de alimentação nominal. A operação a

tensões de alimentação mais baixas conduz a uma redução da dissipação de potência nos circuitos, permitida pelas grandes margens de segurança existentes. Assim, havendo margens de segurança grandes, é possível optimizar a performance (podendo trabalhar a frequências mais elevadas) ou optimizar a dissipação de potência (podendo trabalhar a tensões de alimentação mais baixas, reduzindo a dissipação de potência).

Durante algumas etapas do projecto e fabricação do circuito integrado, é possível reduzir as margens de erro iniciais, uma vez que à medida que a implementação física vai sendo realizada, alguma variabilidade vai sendo estabilizada. No entanto, algumas margens de segurança são sempre necessárias para evitar erros inesperados, sobretudo em aplicações onde a segurança é crítica.

2 - Estado da Técnica

O estudo do envelhecimento em circuitos integrados tem sido amplamente estudado nos últimos anos [23][24][42]. Os efeitos de NBTI e PBTI sobre os circuitos têm sido também amplamente simulados e já foram apresentados modelos matemáticos que prevêem a degradação dos circuitos, tanto para operação estática como dinâmica [25][26][35][41]. Foram também já apresentadas na literatura algumas soluções para melhorar a tolerância dos circuitos a degradações da tensão limiar de condução (Vth) [20][27]. Estes métodos usam técnicas adaptativas, polarizando o substrato dos transístores e ajustando a tensão de alimentação do circuito ao longo do tempo, para compensar as variações provocadas por variações de processo de fabrico ou de envelhecimento. Esta solução permite optimizar a performance do circuito mas não

impede a ocorrência de erros. Além disso, o ajuste dos parâmetros não é feito com base em medidas obtidas em tempo real, mas sim a partir de valores de uma tabela, onde estão guardados os ajustes a fazer em determinados instantes de tempo, calculados antecipadamente por simulação. Em [24] é ainda mostrado como a optimização do projecto pode fazer restringir os efeitos da degradação de performance causada por NBTI.

Outro método inclui o uso de circuitos detectores de erros, em alguns casos auxiliados por lógica adicional para recuperar o estado do sistema após a ocorrência do erro [28][29][30][39][40]. No entanto, embora estes métodos possam corrigir alguns erros verificados nos elementos de memória síncronos (como os flip-flop), a correcção da captura dos dados atrasados pelo flip-flop nem sempre é uma tarefa de fácil implementação nos circuitos, pois pode até requerer outras acções correctivas mais complexas, ou não ser possível a correcção de dados, como acontece em aplicações onde a segurança é crítica.

Foram também já apresentados e estudados sensores globais para circuitos, que detectam identificam genericamente os efeitos do envelhecimento e detectam, globalmente, a degradação de performance existente no circuito [19][20][21][22]. No entanto, estes sensores são globais ao circuito, pelo que não monitorizam a degradação exactamente onde os erros de performance são provocados, que é nas células de memória síncronas (por exemplo os flip-flops) que capturam os caminhos combinatórios que possuem maior tempo de atraso (os caminhos críticos). Para além disso, os sensores globais apresentam uma degradação ao longo da vida diferente do circuito que monitorizam. Por esse motivo, ou (i) são utilizadas margens de segurança pequenas e o circuito pode falhar,

porque um caminho crítico do circuito apresenta uma degradação maior que o próprio sensor está a prever, ou (ii) são utilizadas margens de segurança grandes e o circuito não está optimizado.

Uma abordagem diferente de sensores é a técnica de predição de falha (*circuit failure prediction*) proposta em [31]. A ideia subjacente é a de antecipar a ocorrência de falhas no circuito e sinalizar situações potencialmente críticas antes das falhas começarem a ocorrer. Para realizar esse objectivo, em [31] é utilizado uma captura antecipada do sinal de dados em células de memória previamente seleccionadas. O seu principal objectivo era o de permitir reduzir a margem de erro temporal que é utilizada no sincronismo dos circuitos (definido como *time slack*, ou margem de tempo restante no período do relógio) para acomodar as variações PVTA que possam ocorrer, e que limita significativamente a sua performance. Como solução, apresenta um sensor de envelhecimento (também conhecido por sensor local) que fica acoplado aos flip-flops que terminam os caminhos críticos, e que permite detectar a perda de performance pela definição de um atraso máximo de sinal para esse caminho. É ainda de notar que esta solução tem a vantagem, relativamente aos sensores globais, de prever a degradação de performance localmente, isto é, onde o erro de sincronismo ocorre, ou seja, nos flip-flops.

Posteriormente estes conceitos foram melhorados e foi apresentada uma metodologia de sensores de envelhecimento, focando uma nova aplicação dos sensores e com um circuito melhorado [11][12][32]. A metodologia visa aplicações não tolerantes a falhas e pressupõe a monitorização do circuito durante o seu tempo de vida útil, quando este está sujeito a variações de tensão de

alimentação e de temperatura. Inclui um novo sensor de envelhecimento, com uma banda de guarda (tempo de observação dos sinais de dados críticos) programável, resiliente ao envelhecimento provocado por NBTI e com pouca sensibilidade a variações do processo de fabrico, de tensão ou de temperatura. A metodologia inclui também o estudo do processo de monitorização do chip, incluindo a inserção automática de sensores. Mais uma vez esta é também uma solução de sensores locais preditivos.

Porém, as soluções para os sensores de envelhecimento e de performance locais até aqui relatadas têm algumas desvantagens associadas, nomeadamente: (i) os circuitos dos sensores têm que ter uma performance superior e ser menos sensíveis a variações de processo de fabrico, de tensão de alimentação e de temperatura (PVTA), do que o circuito que vão monitorizar (CUT, Circuit Under Test); (ii) as soluções de sensores locais para monitorizar a degradação de performance utilizam um intervalo de tempo (Banda de Guarda) para fazer a monitorização, tempo esse que tem que ser síncrono com o relógio do sistema e ter um período estável perante variações de processo de fabrico, tensão de alimentação ou temperatura; (iii) os sensores apenas podem estar activos em curtos períodos de tempo, para evitar o seu envelhecimento e degradação da sensibilidade do sensor, e durante esses curtos períodos os caminhos críticos podem não ser activados e, consequentemente, o sensor não realiza a sua função; (iv) como a monitorização do sinal de dados é realizada no terminal de entrada do flip-flop, a introdução de sensores afecta as características temporais do caminho, reduzindo a performance do circuito; (v) o sinal que define o tempo de guarda utilizado para sinalizar as transições tardias na entrada do flip-flop tem de ser encaminhado para todos os sensores no circuito e ser

síncrono com o relógio, o que implica o seu tratamento como um sinal de relógio adicional.

Para solucionar os problemas dos sensores locais preditivos existentes, foi ainda apresentada uma solução melhorada para sensores locais de performance [6][13][43] que realiza uma monitorização constante do desempenho do circuito, quando sujeito às variações PVTA. Nomeadamente, apresenta um funcionamento diferente que permite analisar durante o semi-ciclo activo do relógio as transições de dados tardias ocorridas durante a banda de guarda existente no semi-ciclo inactivo anterior do sinal de relógio, e apresenta uma estrutura interna mais simples, permitindo a implementação total do sensor localmente no flip-flop, onde se inclui a geração local da banda de guarda, e dispensando a utilização de um Latch adicional para reter as activações na saída do sensor. As vantagens principais são: (i) o sensor adapta-se às variações dinâmicas VTA, aumentando a sua capacidade de prever a ocorrência de erros quando as variações aumentam; (ii) o sensor realiza uma constante monitorização, estando sempre activo; (iii) a utilização do sensor introduz um baixo aumento de área ao circuito, uma degradação de performance desprezável, e a sua utilização necessita de um mecanismo de controlo simples e local.

Contudo, embora os sensores locais tenham melhorado muito a sua eficácia e simplicidade de utilização, estes sensores só detectam atrasos de propagação anormais se os caminhos críticos que monitorizam forem activados, em modo funcional. Já os sensores globais, embora possam permitir uma constante activação e monitorização dos parâmetros a medir, não monitorizam os erros onde eles ocorrem, podendo conduzir a circuitos não optimizados.

A nova solução proposta nesta invenção propõe utilizar os 2 tipos de sensores atrás mencionados, globais e locais, para permitir a previsão de erros de performance de uma maneira mais eficaz, que possibilite a activação de mecanismos que impeçam a ocorrência de erros durante toda a vida útil de um circuito, permitindo optimizar constantemente o seu funcionamento. Assim é possível desenvolver circuitos que funcionem no limite das suas capacidades temporais e sem falhas, evitando a utilização de margens de erro elevadas para admitir as variações de performance provocadas por variações no processo de fabrico, na tensão de alimentação, na temperatura ou o envelhecimento. Propõe-se ainda o sistema de controlo que permite, depois da detecção de um potencial erro, desencadear um processo para diminuir a frequência do sinal de relógio do sistema, ou aumentar a tensão de alimentação, evitando que o erro ocorra.

No passado outras técnicas para controlo dinâmico da operação de circuitos integrados já foram apresentadas. Destacam-se técnicas de DVS (*Dynamic Voltage Scaling*) [38], de DFS (Dynamic Frequency Scaling) [45], ou ambas (DVFS – Dynamic Voltage and Frequency Scaling) [44]. Das técnicas existentes, os mecanismos de controlo que permitem alterar a tensão de alimentação e/ou a frequência baseiam-se em taxa de erros existente, ou em estatística de operação efectuada (histórico), ou em tabelas de consulta, ou em sensores globais, ou em casos mais radicais em erros corrigidos em sensores locais, ou em soluções mistas. Porém, as soluções existentes ou são de muito complexa implementação, ou apresentam margens de segurança elevadas, conduzindo a soluções em que a operação do circuito não está optimizada.

A solução apresentada nesta invenção, de utilizar sensores preditivos locais e globais que são sensíveis ao envelhecimento a longo prazo ocorrido nos circuitos, constitui uma novidade no estado da técnica relativamente ao controlo de sistemas de DVS e/ou DFS.

3 - Sumário da invenção

Nesta invenção propõe-se um método de ajuste da tensão de alimentação ou da frequência, que é sensível ao envelhecimento a longo prazo dos circuitos e que utiliza sensores globais e locais. A optimização da performance dos circuitos é alcançada através do aumento da frequência de operação até ao limite máximo que ainda previne a ocorrência de erros. A optimização de consumo de energia é obtida através da redução da tensão de alimentação (VDD) para o valor mínimo que ainda evita a ocorrência de erros. Os sensores globais realizam a monitorização periódica dos atrasos, ou quando solicitado, enquanto que os sensores locais prevêem a ocorrência de erros localmente, nos elementos de memória que apresentam maior probabilidade de ocorrência de erros (células de memória críticas). Tanto os sensores globais como os locais podem desencadear ajustes na frequência ou na tensão de alimentação. O projecto dos sensores e a inserção dos sensores locais no circuito deve ser realizada através de uma análise de previsão de envelhecimento, para determinar os caminhos críticos e os caminhos de sinal de envelhecimento mais rápido (que se podem tornar críticos durante a vida útil do circuito).

As novidades apresentadas na presente invenção são de 2 níveis: no mecanismo de controlo que permite a

optimização dinâmica da tensão ou da frequência e na arquitectura e funcionamento do sensor global a inserir no circuito.
Relativamente ao controlo do sistema de optimização dinâmica, existe novidade: (i) na utilização conjunta de sensores locais e globais para garantir níveis de optimização elevados, (ii) na utilização de sensores preditivos (globais e locais) que previnem os erros de ocorrerem e (iii) na utilização de sensores sensíveis ao envelhecimento do circuito ao longo da sua vida útil. O sensor global para monitorização de variações PVTA é também uma novidade (iv), já que apresenta sensores para a degradação nos transístores PMOS e sensores para a degradação nos transístores NMOS.

A presente invenção é seguidamente descrita em pormenor, sem carácter limitativo e a título exemplificativo, por meio de uma forma de realização exemplificativa, representada nos desenhos anexos, nos quais:

- fig.1 é uma representação em diagrama de blocos dos elementos constituintes da invenção.

- fig.2 é uma representação princípio de operação da invenção, materializado na máquina de estados do controlador do sistema de optimização.

- fig.3, é uma representação em diagrama de blocos e esquemático simplificado da unidade de sensores globais

- fig.4 é uma representação em esquemático de um sensor unitário utilizado na unidade de sensores globais.

- fig.5 é uma representação da arquitectura e ligação dos portos de uma porta NOR (a) e de uma porta NAND (b) das cadeias de NORs e NANDs existentes na unidade de sensores globais.

4- Descrição da concretização preferida

Fazendo referência às figuras, vai agora ser descrita a concretização preferida do invento, em que o dispositivo é constituído por um conjunto de elementos representados nas ditas figuras, montados como seguidamente se descreve.

O método descrito na presente invenção (fig. 1) inicia-se com a análise do circuito onde vai ser instalado o sistema de optimização dinâmica. Esta análise deve identificar os caminhos críticos existentes no circuito e, através de uma metodologia de previsão de envelhecimento (como descrita em [13]), os caminhos que previsivelmente irão envelhecer mais e se vão tornar críticos durante a vida útil espectável para o circuito. Uma vez identificados os caminhos críticos, irão ser inserido os sensores locais (3) através da substituição dos flip-flops que terminam os caminhos críticos identificados por novos flip-flops que incluem sensores de performance e de envelhecimento. É de referir que estes sensores estão descritos em [43] e são sensores preditivos, ou seja, que sinalizam precocemente os erros de performance, antes de eles ocorrerem nos flip-flops que capturam os caminhos críticos.
Com o conhecimento (e simulação) sobre os caminhos críticos do circuito, poderão ser definidas duas cadeias de portas, (6) e (7), que criam dois caminhos fictícios,

com tempos de propagação superiores aos caminhos críticos que se esperam vir a ter no circuito durante a sua vida útil. Uma cadeia é implementada com portas NOR (6), criando um caminho crítico fictício (18) que irá, presumivelmente, envelhecer mais que os caminhos críticos do circuito quando sujeitos ao efeito NBTI (que influencia fortemente a degradação do Vth dos transístores PMOS). A outra cadeia é implementada com portas NAND (7), criando um caminho crítico fictício (19) que irá, presumivelmente, envelhecer mais que os caminhos críticos do circuito quando sujeitos ao efeito PBTI (que influencia fortemente a degradação do Vth dos transístores NMOS).

Uma vez definidas as cadeias com os caminhos críticos fictícios, a unidade de sensores globais (5) pode ser completamente definida. É constituída por uma unidade de controlo do sensor global, (8) e (17), que recebe a ordem do controlador do sistema para iniciar a análise ao desempenho do circuito e gera os sinais de controlo (20, 21, 22 e 23) para a operação de análise global do desempenho. Esta operação permite colocar transparentes as entradas das portas NOR e NAND das cadeias, através dos sinais (21) e (23), para que através dos sinais (20) e (22) seja gerada uma sequência de teste que estimule as duas transições de estado nas saídas das portas das 2 cadeias.

Ao longo das duas cadeias, irão ser ligadas diversos sinais à saída de algumas portas NOR e NAND (36, 41), sinais esses que serão conectados a células de sensores globais (13 e 24), criando diversos caminhos fictícios com diferentes tempos de propagação. Estas células de sensores globais, definidas na figura 4 (24), são semelhantes aos sensores locais já mencionados, mas com a funcionalidade de Latch em vez de Flip-flop. Internamente

são, por isso, constituídos por um Latch ligado a um elemento de atraso (26) e a um elemento detector de estabilidade (27). Estes elementos estão definidos já em [43]. As saídas dos sensores das duas cadeias formam duas saídas de dados do sensor global, (15) e (16).

Para que a optimização do desempenho do circuito seja maximizada, deverá ser iniciado em modo de teste a calibração dos sensores, onde irão ser estimulados alguns caminhos críticos no circuito (através de um teste determinístico) e, ao mesmo tempo, iniciada a análise do desempenho pela unidade de sensores globais. Este procedimento, executado repetidamente para diferentes valores da frequência do relógio (ou da tensão de alimentação), permitirá definir qual o limite máximo (mínimo) para frequência (tensão de alimentação) que garante uma operação do circuito sem que os sensores locais sejam sinalizados. Esta informação da frequência (tensão) irá ser guardada num registo do controlador (registo V/F) e corresponde à frequência (tensão) de trabalho. Para além disso, a operação simultânea dos sensores globais permite determinar quais os caminhos fictícios nas duas cadeias apresentam tempos de propagação semelhantes aos caminhos críticos do circuito. Esta informação será também guardada em dois registos no controlador do sistema (registos GSO_{safe}), que indicam o estado das saídas dos controladores globais para a operação correcta (optimizada) do circuito.
Posteriormente, durante a vida útil do circuito, o controlador do sistema de optimização (4) irá proceder ao ajuste automático da frequência (ou da tensão de alimentação) do circuito, controlando a unidade de gestão do relógio (ou a unidade de gestão da tensão de alimentação) (2) presentes no circuito (1). O

funcionamento durante a vida útil do sistema está simplificado na figura 2, através da máquina de estados simplificada do controlador. O estado (9) corresponde ao funcionamento normal do circuito, estando por isso a ocorrer a análise pelos sensores locais. No entanto, como a activação dos sensores locais depende do funcionamento do circuito e da sua utilização, periodicamente é activada a análise pelos sensores globais, transitando o sistema para o estado (10). Caso não seja sinalizada uma operação diferente dos sensores globais (15 e 16) em relação ao que está guardado em memória como operações correctas, não existe qualquer alteração ao desempenho do circuito e o controlador retorna ao estado inicial (9). Caso o resultado da análise dos sensores globais detecte uma alteração em relação à operação correcta em memória, então iniciar-se-á o processo de alteração do desempenho do circuito com o controlador do sistema no estado (11). Este processo irá fazer alterar (aumentar ou diminuir, consoante os resultados da análise global) a frequência de operação (ou tensão de alimentação), alterando o conteúdo do registo que guarda a frequência (tensão) de trabalho.

Se por ventura ocorrer a sinalização de um sensor local durante a operação normal do circuito, irá ser activada a análise pelos sensores globais e em caso de sinalização de alterações, o procedimento segue como já descrito atrás. No caso de não existir nenhuma sinalização para alteração do desempenho pelos sensores globais, mas tiver existido a sinalização num sensor local, quer dizer que o circuito pode ter envelhecido mais que os caminhos fictícios dos sensores globais, pelo que a frequência (tensão de alimentação) de funcionamento deve ser alterada, mas também deve existir uma actualização nos

registos que guardam a saída correcta dos sensores globais.

É de notar que, se os caminhos fictícios envelhecem mais do que o circuito, as margens de segurança (*time slack*) existentes vão sendo aumentadas ao longo da vida do circuito, tratando-se de uma segurança positiva. Aliás, o projecto e implementação dos caminhos fictícios e as ligações efectuadas entre as portas dos caminhos fictícios (utilizando as ligações descritas na figura 5 para as ligações das portas NOR e NAND, ou ainda utilizando portas com mais entradas), permitem definir caminhos fictícios com elevada probabilidade de envelhecimento, de forma a que estes caminhos envelheçam sempre mais que os caminhos do circuito. No entanto, caso ainda assim exista a possibilidade de o envelhecimento ser maior nos caminhos do circuito, a existência dos sensores locais sempre a monitorizar o desempenho do circuito garantem que o sistema pode aprender com as sinalizações pelos sensores locais e vai-se adaptando às novas condições de operação ao longo da vida útil do circuito.

Este método de optimização e as topologias apresentadas podem ser expandidas e aplicadas a outros tipos de flip-flops, ou utilizando outros tipos de sensores, ou outros caminhos fictícios nos sensores globais, sem prejuízo do método global de optimização que conjuga os 2 tipos de sensores, locais e globais, para optimizar a tensão de alimentação e a frequência de operação.

Faro, 27 de Agosto de 2013

Referências

[1] W. Wang, Z. Wei, S. Yang, Y. Cao, "An Efficient Method to Identify Critical Gates Under Circuit Aging", Proc. IEEE/ACM International Conference on Computer-Aided Design (ICCAD'07), NJ, USA, ISBN: 1-4244-1382-6, 2007.

[2] B. C. Paul, K. Kang, H. Kufluoglu, M. A. Alam, K. Roy, "Temporal Performance Degradation Under NBTI: Estimation and Design for Improved Reliability of Nanoscale Circuits", Tran. Computer-Aided Design of Integrated Circuits and Systems, European Design and Automation Association 3001, Sonoma, CA, USA, Vol. 26 Issue: 4, pp. 743-751, DOI: 10.1109/TCAD.2006.884870, 2007.

[3] J. G. Massey, "NBTI: What We Know and What We Need to Know - A Tutorial Addressing the Current Understanding and Challenges for the Future", Integrated Reliability Workshop Final Report (IRW'04), IEEE, pp. 199-211, DOI: 10.1109/IRWS.2004.1422784, 18-21 Oct., 2004.

[4] V. Huard, F. Monsieur, G. Ribes, S. Bruyere, "Evidence for Hydrogen-Related Defects During NBTI Stress in p-MOSFETs", Proc. 41st Annual Reliability Physics Symposium, IEEE International, Dallas, Texas, pp. 178-182, DOI: 10.1109/RELPHY.2003.1197741, 2003.

[5] J. Semião et al., "Time Management for Low-Power Design of Digital Systems", ASP Journal of Low Power Electronics (JOLPE), Vol. 4, N° 3, pp. 410-419, Dec., 2008.

[6] C. V. Martins, J. Semião, J. C. Vazquez, V. Champaq, M. Santos, I. C. Teixeira, J. P. Teixeira, "Adaptive Error-Prediction Flip-flop for Performance Failure

Prediction with Aging Sensors", 29th IEEE VLSI Test Symposium 2011 (VTS'11), Dana Point, California, USA, 1 -5 May, 2011.

[7] H. Abrishami, S. Hatami, B. Amelifard, M. Pedram, "Multi-Corner, Energy-Delay Optimized, NBTI-Aware Flip-Flop Characterization and Design", Proc. 18th ACM Great Lakes Symposium on VLSI, ACM New York, NY, USA, pp. 652-659, DOI: 10.1109/ISQED.2010.5450509, 22-24 March, 2010.

[8] A. Tiwari, J. Torrellas, "Facelift: Hiding and Slowing Down Aging in Multicores", Proc. 41st Annual IEEE/ACM International Symposium on Microarchitecture, IEEE Computer Society Washington, DC, USA, pp. 129-140, ISBN: 978-1-4244-2836-6, 8-12 Nov, 2008.

[9] J. Abella, X. Vera, A. González, "Penelope: The NBTI-Aware Processor", Proc. 40th IEEE/ACM International Symposium on Microarchitecture, IEEE Computer Society Washington, DC, USA, pp. 85-96, DOI: 10.1109/MICRO.2007.11, 1-5 Dec., 2007.

[10] S. Tsujikawa, Y. Akamatsu, H. Umeda, J. Yugami, "Two Concerns about NBTI Issue: Gate Dielectric Scaling and Increasing Gate Current", Proc. 42nd Annual Reliability Physics Symposium (ARPS'04), IEEE, pp. 28-34, DOI: 10.1109/RELPHY.2004.1315297, 25-29 April, 2004.

[11] J. C. Vazquez et al., "Built-In Aging Monitoring for Safety-Critical Applications", Proc. IEEE Int. On-Line Test Symp. (IOLTS), pp. 9-14, 2009.

[12] J. C. Vazquez et al., "Low-sensitivity to Process Variations Aging Sensor for Automotive Safety-Critical Applications", Proc. IEEE VLSI Test Symposium (VTS), pp. 238-243, 2010.

[13] C. V. Martins, J. Pachito, J. Semião, I. C. Teixeira, J. P. Teixeira, "Adaptive Error-Prediction Aging Sensor for On-Line Monitoring of Performance Errors", Proceedings of the XXVI Conference on Design of Circuits and Integrated Systems - DCIS'11, Albufeira, Portugal, November, 2011.

[14] T. K. Maiti, S. S. Mahato, P. Chakraborty, C. K. Maiti, S. K. Sarkar, "Negative Bias Temperature Instability in Strain-Engineered p-MOSFETs: A Simulation Study", Journal of Computational Electronics, Vol. 9, N° 1, pp. 1-7, DOI: 10.1007/s10825-009-0270-6, 2010.

[15] B. Zhu, J. S. Suehle, J. B. Bemstein, Y. Chen, "Mechanism of Dynamic NBTI of pMOSFETs", Integrated Reliability Workshop Final Report (IRW'04), IEEE, pp. 113-117, DOI: 10.1109/IRWS.2004.1422751, 18-21 Oct., 2004.

[16] J. Srinivasan, S. V. Adve, P. Bose, J. A. Rivers, "The Case for Lifetime Reliability-Aware Microprocessors", Proc. 31st Annual International Symposium on Computer Architecture (ISCA'04), IEEE Computer Society Washington, DC, USA, pp. 276-287, ISBN: 0-7695-2143-6, 19-23 June, 2004.

[17] H. Kufluoglu, M. A. Alam, "A Computational Model of NBTI and Hot Carrier Injection Time-Exponents for MOSFET Reliability", Journal of Computational Electronics, Vol. 3, N° 3-4, pp. 165-169, DOI: 10.1007/s10825-004-7038-9, 2004.

[18] J. Wang, W.M. Huang, H. Bor-Yuan, M. Racanelli, J. Foerstner, J. Woo, "Hot-Carrier-Injection (HCI) Immunity Under High Drain Stress of Thin-Film SOI n-MOSFETs Fabricated on SIMOX and BESOI Substrates", Proc. International SOI Conference, IEEE, Nantucket,

MA, USA, pp. 129-130, DOI: 10.1109/SOI.1994.514280, 3-6 Oct., 1994.

[19] D. Kim, J. Kim, M. Kim, J. Moulic, H. Song, "System and Method for Monitoring Reliability of a Digital System", IBM Corp., US Patent 7495519, Feb. 24, 2009.

[20] J. Tschanz, et al. "Adaptive Frequency and Biasing Techniques for Tolerance to Dynamic Temperature-voltage Variations and Aging," Proc. IEEE Int. Solid-State Circ. Conf. (ISSCC), pp. 292-293, 2007.

[21] C. R. Gauthier, P. R. Trivedi, G. S. Yee, "Embedded Integrated Circuit Aging Sensor System", Sun Microsystems, US Patent 7054787, May 30, 2006.

[22] J. Keane, T. Kim, C. Kim, "An on-chip NBTI sensor for measuring PMOS threshold voltage degradation", Proc. Int. Symp. on Low Power Electronics and Design (ISLPED), pp. 189-194, 2007.

[23] Wenping Wang, Shengqi Yang, S. Bhardwaj, R. Vattikonda, S. Vrudhula, T. Liu, Yu Cao, "The Impact of NBTI on the Performance of Combinational and Sequential Circuits", Proc. 44th ACM/IEEE Design Automation Conf. (DAC), pp. 364-369, 2007.

[24] K. Kang, S. Gangwal, S. Phil Park, and K. Roy, "NBTI Induced Performance Degradation in Logic and Memory Circuits: How Effectively Can We Approach a Reliability Solution?", Proc. Asia / South Pacific Design Autom. Conf. (ASP-DAC), pp. 726-731, 2008.

[25] R. Vattikonda, W. Wang, Y. Cao, "Modeling and Minimization of PMOS NBTI Effect for Robust Nanometer Design", Proc. DAC, pp. 1047-1052, 2006.

[26] S. Bhardwaj, W. Wang, R. Vattikonda, Y. Cao, and S. Vrudhula, "Predictive Modeling of the NBTI Effect for Reliable Design," in Proceedings of the IEEE

Custom Integrated Circuits Conference (CICC), September 2006.

[27] S.V. Kumar, C.H. Kim, S. Sapatnekar, "Adaptive Techniques for Overcoming Performance Degradation due to Aging in Digital Circuits", Proc. IEEE ASP-DAC, pp. 284-289, Jan. 2009.

[28] Dan Ernst, Nam Sung Kim, Shidhartha Das, Sanjay Pant, Toan Pham, Rajeev Rao, Conrad Ziesler, David Blaauw, Todd Austin, Trevor Mudge, "Razor: A Low-Power Pipeline Based on Circuit-Level Timing Speculation", Micro Conference, December, 2003.

[29] Dan Ernst, Shidhartha Das, Seokwoo Lee, David Blaauw, Todd Austin, Trevor Mudge, Nam Sung Kim, Krisztián Flautner, "Razor: Circuit-Level Correction of Timing Errors for Low-Power Operation", IEEE Micro, vol. 24, n° 6, pp 10-20, November 2004.

[30] D. Blaauw, S. Kalaiselvan, K. Lai, Wei-Hsiang Ma, S. Pant, C. Tokunaga, S. Das, D. Bull, "Razor II: In Situ Error Detection and Correction for PVT and SER Tolerance," Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International , vol., no., pp.400-622, 3-7 Feb. 2008.

[31] M. Agarwal, et al., “Circuit Failure Prediction and Its Application to Transistor Aging”. Proc. VLSI Test Symp. (VTS), pp. 277-286, 2007.

[32] J.C. Vazquez, et al., “Predictive Error Detection by On-line Aging Monitoring”, Proc. IEEE Int. On-Line Test Symp. (IOLTS), 2010.

[33] Gregor Pobegena, Thomas Aichingerb, Tibor Grasserc, Michael Nelhiebeld, “Impact of gate poly doping and oxide thickness on the N- and PBTI in MOSFETs”, Proceedings of the 22th European Symposium on the “Reliability of Electron Devices, Failure Physics

and Analysis", Microelectronics Reliability, Volume 51, Issues 9–11, September–November 2011, Pages 1530–1534.

[34] B. Cheng, A. R. Brown, S. Roy, and A. Asenov, "PBTI/NBTI-Related Variability in TB-SOI and DG MOSFETs", IEEE Electron Device Letters, Vol. 31, Issue No. 5, pages 408 – 410, ISSN: 0741-3106, May 2010.

[35] K. Joshi, S. Mukhopadhyay, N. Goel, S. Mahapatra, "A consistent physical framework for N and P BTI in HKMG MOSFETs", 2012 IEEE International Reliability Physics Symposium (IRPS), Pages: 5A.3.1 - 5A.3.10, DOI: 10.1109/IRPS.2012.6241840, 2012.

[36] M. Sato, N. Umezawa, J. Shimokawa, H. Arimura, S. Sugino, A. Tachibana, M. Nakamura, N. Mise, S. Kamiyama, T. Morooka, T. Eimori, K. Shiraishi, K. Yamabe, H. Watanabe, K.Yamada, T. Aoyama, T. Nabatame, Y. Nara, and Y. Ohji, "Physical Model of the PBTI and TDDB of La Incorporated HfSiON Gate Dielectrics with Pre-existing and Stress-induced Defects", IEEE International Electron Devices Meeting (IEDM), 2008, DOI: 10.1109/IEDM.2008.4796629.

[37] Asen Asenov, Andrew R. Brown, Binjie Cheng, "Statistical aspects of NBTI/PBTI and impact on SRAM yield", Design, Automation & Test in Europe Conference & Exhibition (DATE), DOI: 10.1109/DATE.2011.5763240, 2011.

[38] S. Das, D. Roberts, S. Lee, S. Pant, D. Blaauw, T. Austin, T. Mudge, K. Flautner, "A self-tuning DVS Processor using Delay Error Detection and Correction", IEEE J. Solid-Sate Circuits, vol. 41, n° 4, pp. 792-804, April 2006.

[39] Yang Lin, Mark Zwolinski, “SETTOFF : A Fault Tolerant Flip-Flop for Building Cost-efficient Reliable Systems”, IEEE 18th International On-Line Testing Symposium (IOLTS), June 2012, Page(s): 7 – 12, DOI: 10.1109/IOLTS.2012.6313833

[40] Martin Omaña, Daniele Rossi, Nicolò Bosio, Cecilia Metra, “Low Cost NBTI Degradation Detection and Masking Approaches”, IEEE Transactions on Computers, Vol. 62, Issue No. 3, DOI: 10.1109/TC.2011.246, March 2013.

[41] M. A. Alam, S. Mahapatra, "A Comprehensive Model of PMOS NBTI Degradation", Journal of Microelectronics Reliability, Vol. 45 Issue: 1, pp. 71-81, DOI: 10.1016/j.microrel.2004.03.019, January, 2005.

[42] W. Wang, S. Yang, S. Bhardwaj, S. Vrudhula, F. Liu, Y. Cao, "The Impact of NBTI Effect on Combinational Circuit: Modeling, Simulation, and Analysis", Tran. Very Large Scale Integration (VLSI'10), IEEE, Vol. 18 Issue: 2, pp. 173-183, DOI: 10.1109/TVLSI.2008. 2008810, Feb., 2010.

[43] C. V. Martins, J. Semião, I. C. Teixeira, J. P. Teixeira, "Flip-Flop com Sensor para Previsão de Erros de Performance, para Aplicação como Sensor Local de Performance ou de Envelhecimento em Circuitos Integrados Digitais Síncronos", Patente Portuguesa - PT 105636, Universidade do Algarve, 20 de Abril de 2011.

[44] Joonho Kong, Jinhang Choi, Lynn Choi, Sung Woo Chung, “Low-cost Application-aware DVFS for Multi-core Architecture”, Third 2008 International Conference on Convergence and Hybrid Information Technology (ICCIT’08), Volume 2, Page(s):106 – 111, Nov. 2008, Busan – South Korea, DOI: 10.1109/ICCIT.2008.124

[45] T.R. da Rosa, V. Larrea, N. Calazans, F.G. Moraes, "Power consumption reduction in MPSoCs through DFS", 25th Symposium on Integrated Circuits and Systems Design (SBCCI), 2012, Page(s):1 – 6, Brasilia – Brasil, DOI: 10.1109/SBCCI.2012.6344429

Faro, 17 de Abril de 2012

REIVINDICAÇÕES

1 – Sistema para optimização ao longo da vida da tensão de alimentação e/ou da frequência de operação de circuitos integrados digitais síncronos, caracterizado por um bloco controlador (1), por um bloco de sensores globais (2) e por sensores locais inseridos no circuito digital (3).

2 – Sistema para optimização, de acordo com a reivindicação nº 1, caracterizado por utilizar sensores preditivos que permitem a operação do circuito e a sinalização de situações de erro eminente, sem que os erros efectivamente ocorram.

3 – Sistema para optimização, de acordo com a reivindicação nº 1, caracterizado por utilizar sensores sensíveis ao envelhecimento a longo prazo dos circuitos, permitindo sinalizar situações de erro eminentes ao longo de toda a vida útil do circuito.

4 – Sistema para optimização, de acordo com a reivindicação nº 1, caracterizado por utilizar sensores sensíveis ao envelhecimento a longo prazo dos circuitos, permitindo sinalizar situações de erro eminente ao longo de toda a vida útil do circuito.

5 – Unidade de sensores globais para monitorização da performance em circuitos digitais CMOS, caracterizado por incluir um controlador, uma cadeia de portas NOR, uma cadeia de portas NAND e vários sensores de performance ligados ao longo das duas cadeias.

6 - Unidade de sensores globais para monitorização da performance, de acordo com a reivindicação nº 5, caracterizado por possuir um conjunto de portas NOR ligadas em cadeia, de forma a que a saída da porta anterior na cadeia seja ligada à entrada, da porta NOR seguinte na cadeia, que está ligada ao transístor PMOS que tem o seu terminal de fonte ligado à tensão de alimentação, e que as entradas (de todas as portas) ligadas ao transístor PMOS que não tem o seu terminal de fonte ligado à alimentação estejam todas ligadas a um mesmo sinal de controlo, para permitir ligar as portas todas em cadeia ou isolar as portas e permitir o envelhecimento dos transístores PMOS ligados à alimentação.

7 - Unidade de sensores globais para monitorização da performance, de acordo com a reivindicação nº 5, caracterizado por possuir um conjunto de portas NAND ligadas em cadeia, de forma a que a saída da porta anterior na cadeia seja ligada à entrada, da porta NAND seguinte na cadeia, que está ligada ao transístor NMOS que tem o seu terminal de fonte ligado à massa, e que as entradas (de todas as portas) ligadas ao transístor NMOS que não tem o seu terminal de fonte ligado à massa estejam todas ligadas a um mesmo sinal de controlo, para permitir ligar as portas todas em cadeia ou isolar as portas e permitir o envelhecimento dos transístores NMOS ligados à massa.

8 - Unidade de sensores globais para monitorização da performance, de acordo com a reivindicação nº 5, caracterizado por possuir um sensor de performance ligado a diversas saídas ao longo das cadeias de portas NOR e NAND que inclui uma porta Latch, um elemento de atraso e

um detector de estabilidade com lógica para retenção do sinal de saída em modo activo.

9 - Unidade de sensores globais para monitorização da performance, de acordo com a reivindicação nº 5, caracterizado por possuir um controlador que, uma vez activado por um sinal de entrada, irá desencadear uma sequência de teste para as cadeias de portas NOR e NAND que contenha as duas transições de estado, ou seja, de nível baixo para nível alto, e de nível alto para nível baixo.

10 – Processo de funcionamento do controlador do sistema para optimização ao longo da vida da tensão de alimentação e/ou da frequência de operação de circuitos integrados digitais síncronos, caracterizado por monitorizar constantemente através dos sensores locais as transições de dados tardias nas entradas de dados das células de memória críticas que incluem funcionalidades de sensores de performance, por sinalizar a ocorrência destas transições tardias, e por desencadear um processo de análise de desempenho utilizando a unidade de sensores globais caso essa sinalização ocorra.

11 – Processo de funcionamento do controlador do sistema para optimização ao longo da vida, de acordo com a reivindicação nº 10, caracterizado por desencadear periodicamente a análise de desempenho do circuito pela unidade de sensores globais, mesmo quando não existem sinalizações nos sensores locais.

12 – Processo de funcionamento do controlador do sistema para optimização ao longo da vida, de acordo com a reivindicação nº 10, caracterizado por desencadear a

alteração da frequência de operação ou da tensão de alimentação do circuito caso exista uma sinalização no sensor global que indique que o desempenho das portas lógicas presentes nos caminhos fictícios de propagação são diferentes do esperado e guardado em memória, subindo ou baixando a frequência e a tensão, de acordo com a análise obtida e de forma a que o circuito mantenha um funcionamento sem erros.

13 – Processo de funcionamento do controlador do sistema para optimização ao longo da vida, de acordo com a reivindicação nº 10, caracterizado por desencadear uma actualização da frequência de operação ou da tensão de alimentação e por actualizar em memória os dados que indicam as saídas dos sensores globais que indicam a operação correcta do circuito, caso exista uma sinalização nos sensores locais e não exista uma sinalização nos sensores globais para haver alterações de desempenho..

14 – Processo de funcionamento do controlador do sistema para optimização ao longo da vida, de acordo com a reivindicação nº 10, caracterizado desencadear a activação de alguns caminhos críticos em modo de teste, realizando a activação dos sensores globais e realizando o teste a diferentes tensões ou frequências, para calibração dos sensores globais e definição de qual frequência e tensão de alimentação correcta que permite a operação optimizada do circuito, e para definição das saídas dos sensores globais para essa operação optimizada.

Faro, 27 de Agosto de 2013

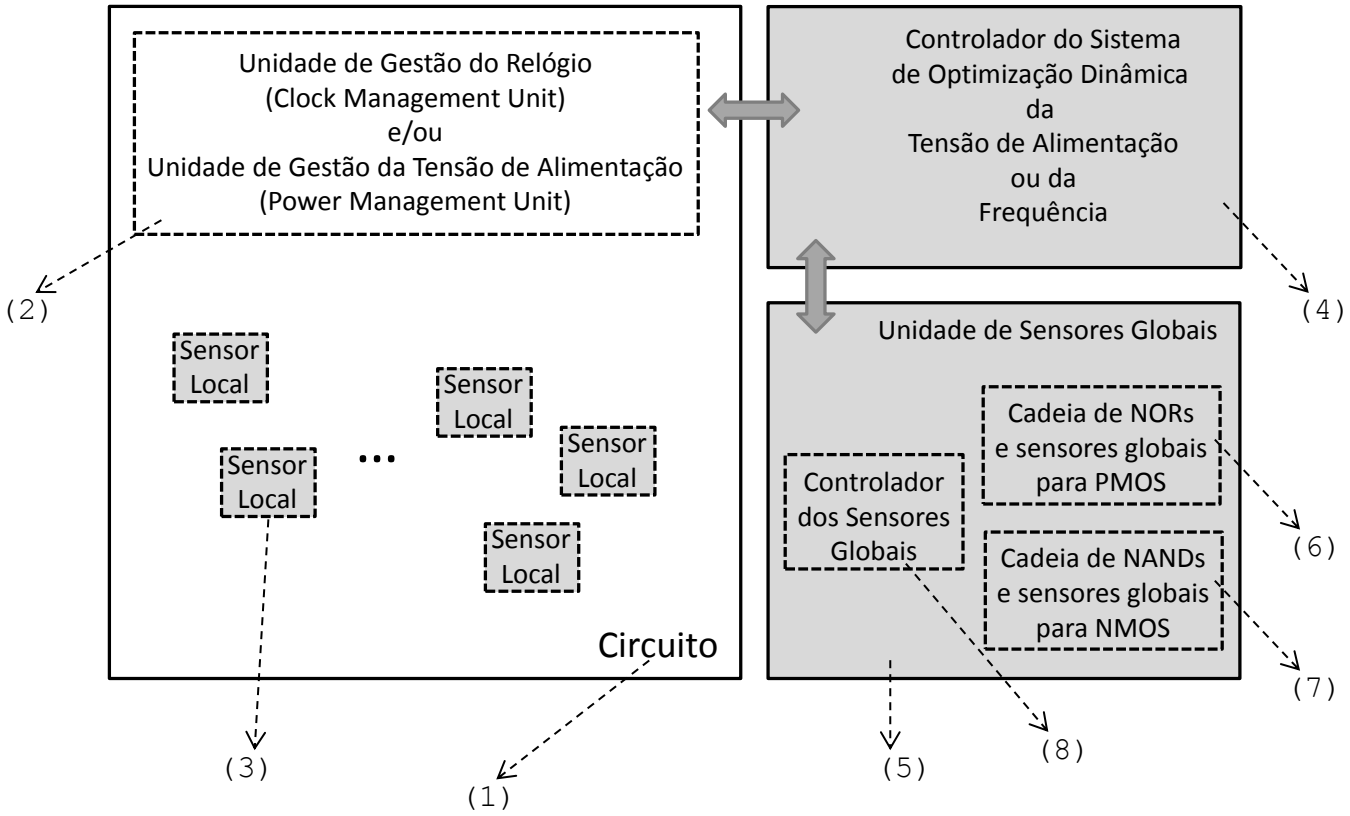

Unidade de Gestão do Relógio
(Clock Management Unit)
e/ou
Unidade de Gestão da Tensão de Alimentação
(Power Management Unit)
Controlador do Sistema
de Optimização Dinâmica
da
Tensão de Alimentação
ou da
Frequência
Sensor Local
Sensor Local
Sensor Local
Sensor Local
Sensor Local
Circuito
Unidade de Sensores Globais
Cadeia de NORs e sensores globais para PMOS
Controlador dos Sensores Globais
Cadeia de NANDs e sensores globais para NMOS
(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)

Fig.1

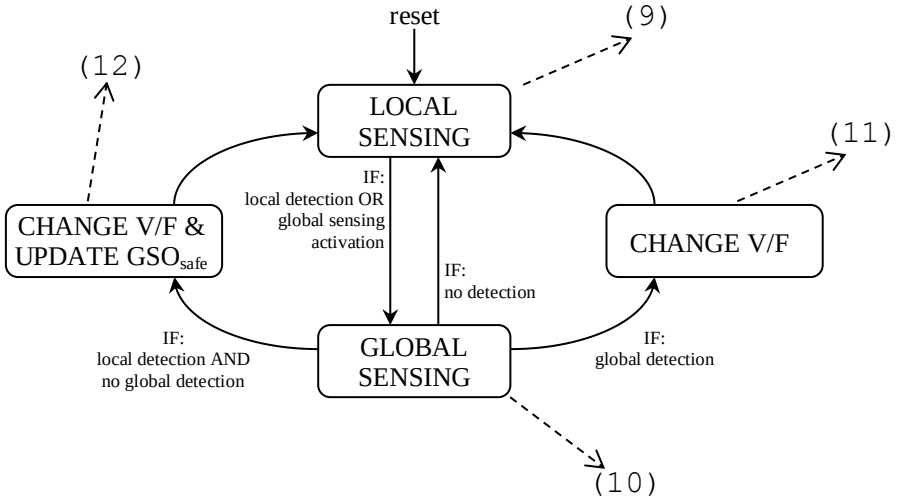

reset
(9)
(12)
LOCAL SENSING
(11)
IF: local detection OR global sensing activation
CHANGE V/F & UPDATE GSO$_{safe}$
CHANGE V/F
IF: no detection
IF: local detection AND no global detection
GLOBAL SENSING
IF: global detection
(10)

Fig.2

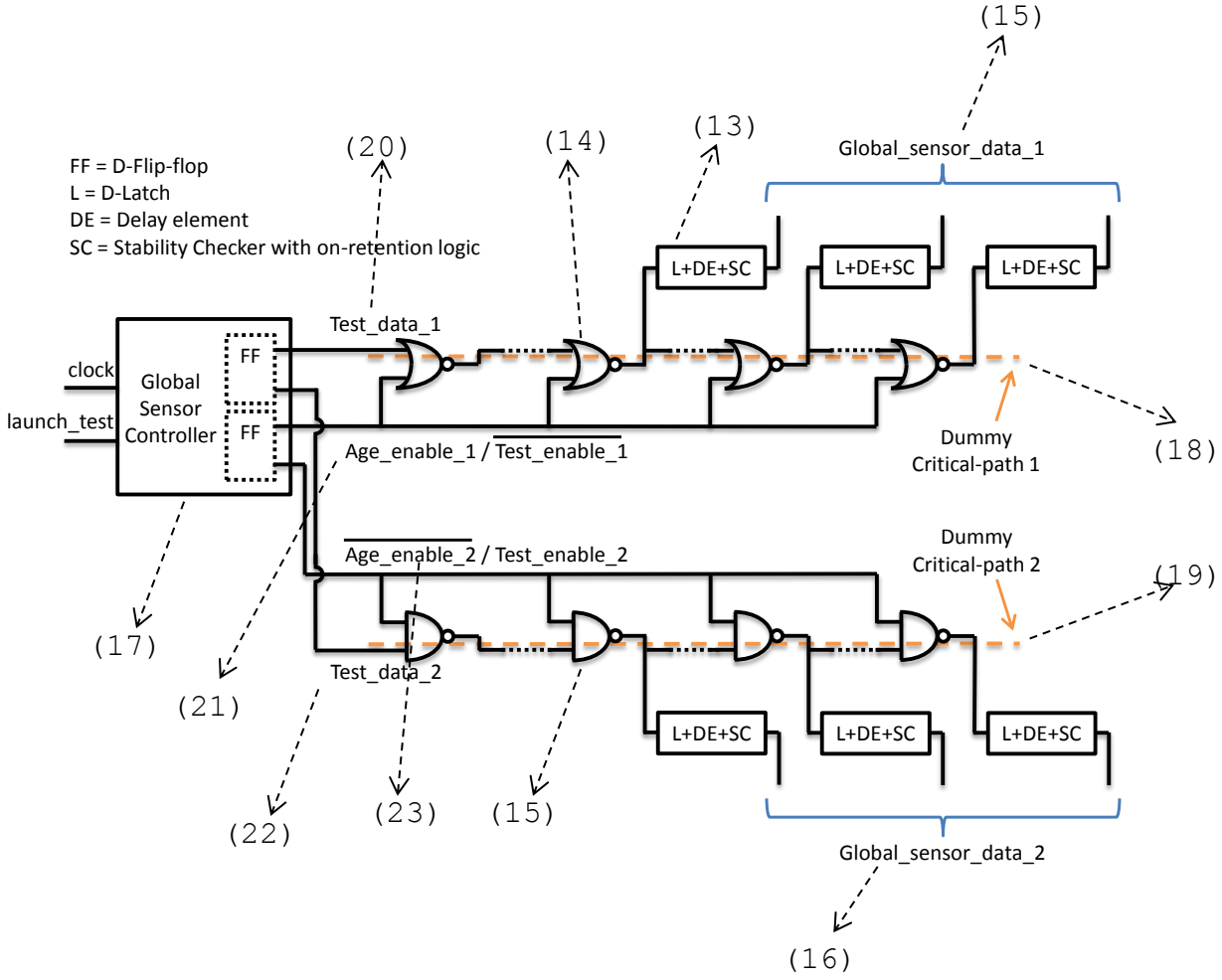

(15)
FF = D-Flip-flop
L = D-Latch
DE = Delay element
SC = Stability Checker with on-retention logic
(20)
(14)
(13)
Global_sensor_data_1
L+DE+SC
L+DE+SC
L+DE+SC
Test_data_1
clock
launch_test
Global Sensor Controller
FF
FF
Age_enable_1 / Test_enable_1
Dummy Critical-path 1
(18)
Age_enable_2 / Test_enable_2
Dummy Critical-path 2
(19)
(17)
(21)
Test_data_2
L+DE+SC
L+DE+SC
L+DE+SC
(22)
(23)
(15)
Global_sensor_data_2
(16)

Fig. 3

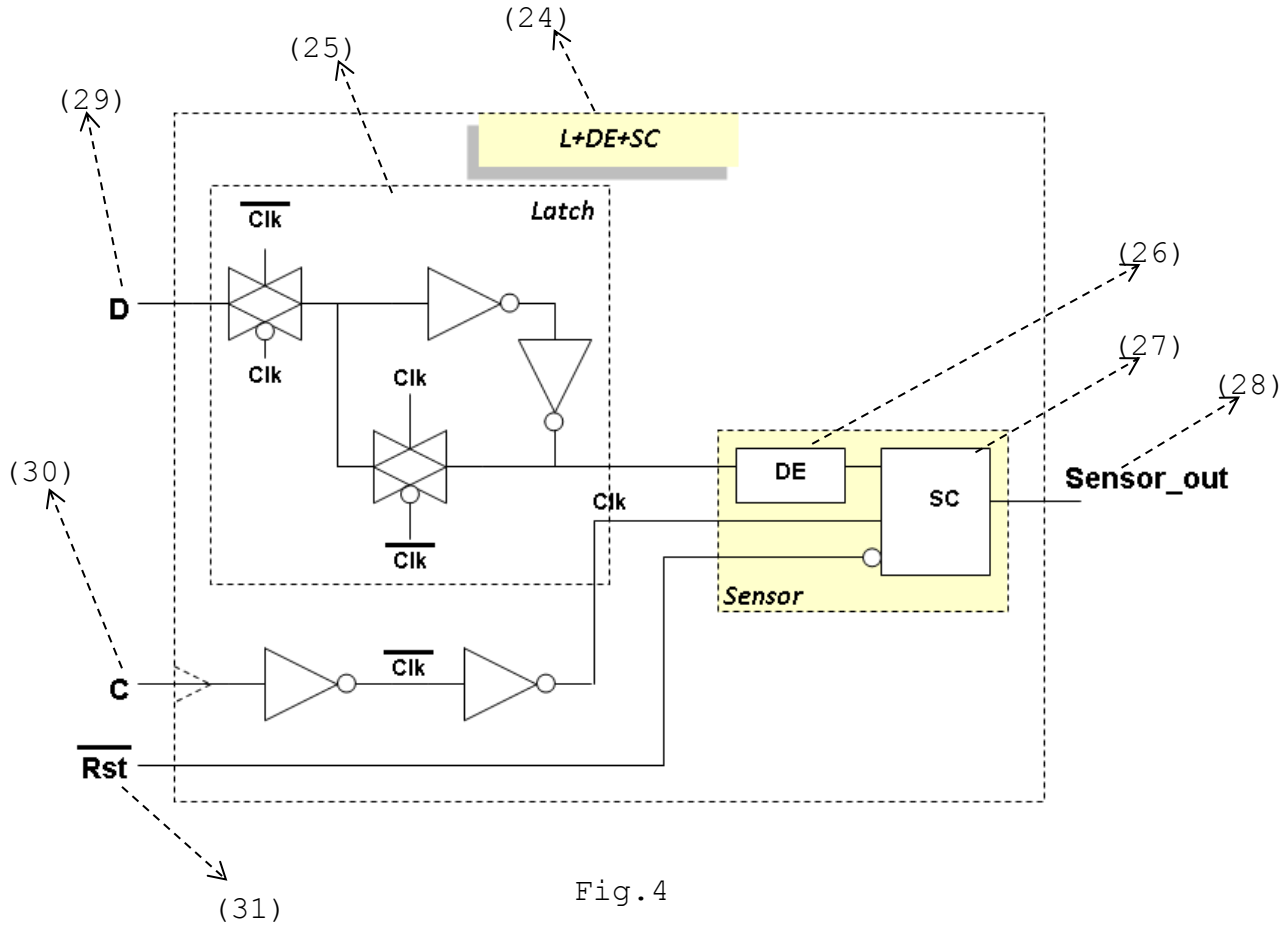
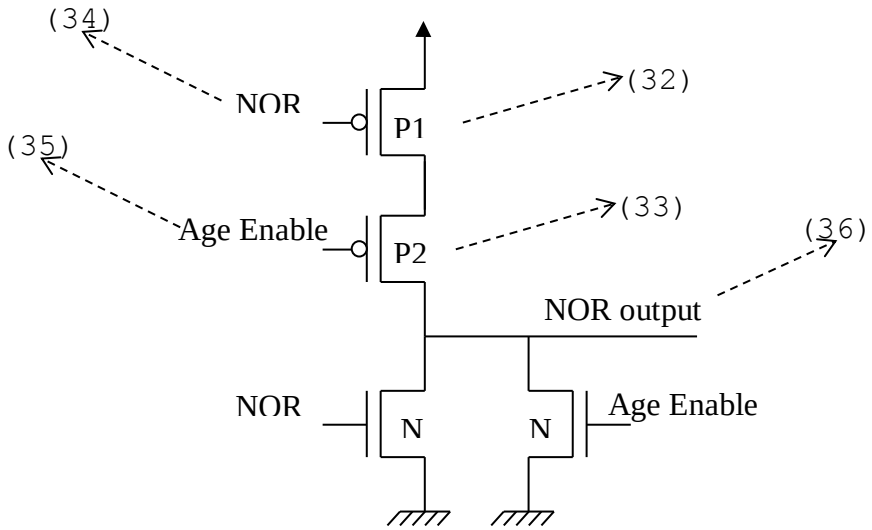
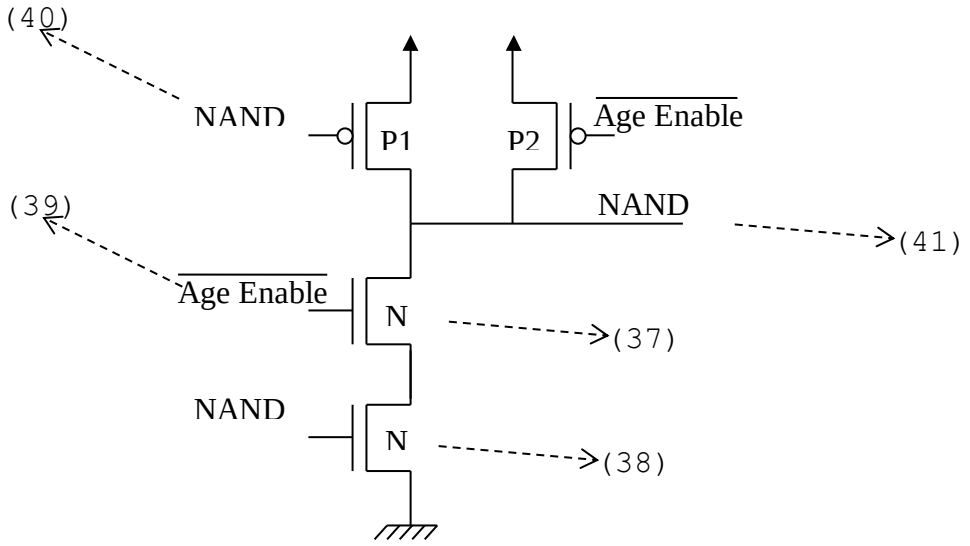

(24)
(25)
(29)
L+DE+SC
Latch
Clk
D
Clk
Clk
(26)
(27)
(28)
DE
SC
Sensor_out
(30)
Clk
Clk
Sensor
C
Clk
Rst
(31)

Fig.4


(34)
NOR
(32)
P1
(35)
Age Enable
(33)
P2
(36)
NOR output
NOR
N
N
Age Enable

(a)


(40)
NAND
P1
P2
Age Enable
(39)
NAND
(41)
Age Enable
N
(37)
NAND
N
(38)

(b)

Fig.5