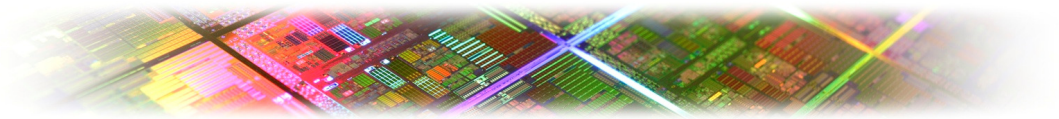
UNIVERSIDADE DO ALGARVE
INSTITUTO SUPERIOR DE ENGENHARIA



ADAPTIVE ERROR-PREDICTION AGING SENSOR FOR SYNCHRONOUS DIGITAL CIRCUITS

SENSOR DE ENVELHECIMENTO COM PREVISÃO DE ERROS ADAPTATIVA PARA CIRCUITOS DIGITAIS SÍNCRONOS

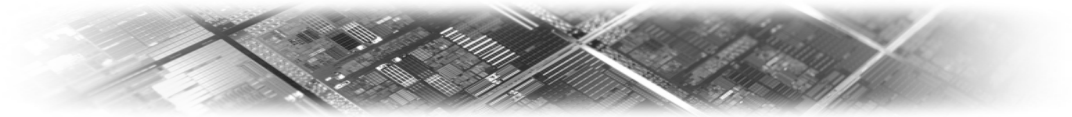
Celestino Virtudes Dias Martins

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

Tutor: Professor Doutor Jorge Filipe Leal Costa Semião

October, 2012

UNIVERSIDADE DO ALGARVE
INSTITUTO SUPERIOR DE ENGENHARIA



ADAPTIVE ERROR-PREDICTION AGING SENSOR FOR SYNCHRONOUS DIGITAL CIRCUITS

SENSOR DE ENVELHECIMENTO COM PREVISÃO DE ERROS ADAPTATIVA PARA CIRCUITOS DIGITAIS SÍNCRONOS

Celestino Virtudes Dias Martins

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

Tutor: Professor Doutor Jorge Filipe Leal Costa Semião

October, 2012

Title: Adaptive Error-Prediction Aging Sensor for Synchronous Digital Circuits

Authorship: Celestino Virtudes Dias Martins

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.

_____

Celestino Virtudes Dias Martins

To my parents.

ACKNOWLEDGEMENTS

I would like to express my sincere thankfulness to Dr. Jorge Semião, my tutor, for all his support and guidance in the elaboration of this dissertation. He is a very generous person, who shared his time, knowledge and resources to help me take this job forward and complete it. The research expertise field that supports my dissertation was completely new to me at the beginning of my Master degree, and my tutor's insightfulness on this field was the reason for me to explore this research area, giving me great confidence for the development of this work.

During the development of this dissertation it was sometimes necessary to abstract myself from my professional work to focus on this task. Hereby I would like to thanks to my colleagues and associates for the patience they had and to the flexibility they provided for me to pursue this goal.

To my parents, Maria Feliciana Dias, my mother, and Manuel Celestino Martins, my father, who always believed in me, I want to thanks for the good advices, on my personal, academic and professional path. Thanks for raising me, I love you both.

To Cátia, a great friend, thanks for the time, the happiness, for sharing and for the sharp advices.

To all my friends, thanks for the good break times that we shared and for boosting me morally. Your presence was invaluable.

Celestino Virtudes Dias Martins, Faro, 29 de Outubro de 2012

ABSTRACT

This paper presents a new approach on aging sensors for synchronous digital circuits. An Adaptive Error-Prediction Flip-Flop (AEP-FF) architecture with built-in aging sensor is proposed, to perform on-line monitoring of long-term performance degradation of CMOS digital systems, regardless of their origin.

The main advantage is that the sensor's performance degradation works in favor of the predictive error detection. Moreover, the sensor is out of the signal path, which allows sensor insertion with negligible performance penalty to the circuit.

Performance error prediction is implemented by the detection of late transitions at flip-flop data input, caused by aging (namely, due to NBTI), by physical defects activated by long lifetime operation, or by other low drift operation dependent performance degradations. Performance errors must not occur in safety-critical systems (automotive, health, space), which highlights the importance of the Performance Failure Prediction (PFP) methodology.

Sensors can be always active, to enhanced Predictive Fault-Detection (PFD), without reducing its sensing capability (in fact, aging degradations over the sensor itself enhances its sensitivity to PFD).

Extensive SPICE simulations were performed with a 65 nm CMOS technology, which uses Berkeley Predictive Technology Models (PTM), to characterize in detail the sensor, and to compare it with previously developed solutions. Simulations show sensor advantages over the existing aging sensors and sensor insertion and operation is validated with two benchmark circuits. It is shown that the impact of aging degradation and/or PVT (Process, power supply Voltage and Temperature) variations on the sensor enhance error prediction.

KEYWORDS: Aging Sensor, NBTI, Performance Failure Prediction, Predictive Fault-Detection.

RESUMO

Este trabalho introduz uma nova abordagem aos sensores de envelhecimento para circuitos digitais síncronos. É proposta uma nova arquitectura para um flip-flop com previsão de erros adaptativa (AEP-FF) para realizar a monitorização *on-line* da perda de performance a longo prazo de sistemas digitais CMOS, independentemente da sua causa.

O AEP-FF integra um FF do tipo TG-MSFF (*Transmission Gate Master Slave Flip-Flop*) e um sensor de envelhecimento composto por um elemento de atraso (DE) ligado a um analisador de estabilidade (SC). A entrada do DE está ligada no nó de saída da *master latch* do FF, e a entrada do SC está ligada na saída do DE. A saída do SC é a saída do sensor do envelhecimento. Durante o semi-ciclo negativo, qualquer sinal na entrada de dados do flip-flop é transmitido para a saída da *master latch* e atrasado pelo DE. Durante este período, o SC ignora quaisquer sinais na sua entrada e mantém a sua saída constante. Quando o sinal de relógio comuta para o semi-ciclo positivo, a porta de transmissão da *master latch* isola a entrada do FF, e a saída do FF é actualizada. No semi-ciclo positivo, o SC fica activo e qualquer variação que ocorra na sua entrada irá ser sinalizada na saída do sensor de envelhecimento. Ou seja, se o sinal na entrada do FF mudar de valor no instante anterior ao flanco ascendente de relógio, essa variação pode propagar-se para o semi-ciclo positivo do sinal de relógio e ser detectada pelo SC, dependendo do momento em ocorra e do tempo de atraso do DE. O SC integra uma *latch* para manter a sua saída activa indefinidamente após uma primeira detecção. Uma entrada de *reset*, no SC, permite repor a saída no nível lógico baixo. Como flip-flop, AEP-FF comporta-se como um flip-flop tipo D *positive-edge trigger*, actualizando a sua saída no flanco ascendente do sinal de relógio.

A previsão de erros de performance consegue-se pela detecção de atrasos anormais do sinal na entrada de dados do flip-flop, provocados por envelhecimento (nomeadamente por NBTI), por variações de processo, tensão de alimentação ou temperatura (PVT), por defeitos físicos latentes, ou por quaisquer outras causas que provoquem a perda gradual de performance do circuito sob teste, antes que esses atrasos possam causar algum dano na operação do circuito. Em última instancia, o

atraso na propagação de sinal, num bloco combinatório de um circuito, pode tornar-se suficientemente alto para impedir que o tempo de setup do flip-flop que termina esse caminho seja garantido, causando uma falha na captura do valor e consequente erro de funcionamento. Esta situação é altamente prejudicial nos circuitos digitais síncronos, uma classe que representa a maioria dos dispositivos electrónicos disponíveis no mercado.

A previsão de erros de performance representa uma solução potencial para os problemas de fiabilidade e de performance nas nanotecnologias, onde os procedimentos tradicionais como o *burn-in* ou a inclusão de bandas de guarda fixas se tornam impraticáveis e/ou muito dispendiosos. A variabilidade, um problema tipicamente reservado aos circuitos analógicos, 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.

O conceito de Previsão de Erros de Performance (PFP) é de extrema importância em sistemas críticos (automóvel, saúde, aeroespacial) onde não podem ocorrer erros. Este conceito, quando associado a mecanismos que permitam o ajuste das condições de funcionamento do circuito (e.g. frequência, tensão de alimentação) às condições de operação do mesmo (temperatura, tensão de alimentação), pode também ser usado para optimizar a performance do circuito, ou o seu consumo de energia, sem comprometer a sua fiabilidade.

Face a outras propostas, uma das principais vantagens do AEP-FF é a introdução de um novo conceito, onde a perda de performance do próprio sensor melhora a sua capacidade de previsão de erros. Os efeitos do envelhecimento, do aumento de temperatura e da diminuição na tensão de alimentação (VTA), aumentam a banda de guarda do sensor e melhoram a sua capacidade de prever erros (PFD), permitindo que o AEP-FF possa estar sempre ligado sem comprometer o seu funcionamento. Outra inovação é introduzida na banda de guarda, que não é definida por um sinal externo síncrono com o sinal de relógio, mas sim através de um atraso extra provocado no

sinal de dados, na entrada do flip-flop. No sensor de envelhecimento, nenhum dos seus elementos tem de ser resiliente a variações de VTA e circuitos simples são usados para implementar o DE e definir a banda de guarda. Além dos circuitos propostos, o DE pode ser implementado por qualquer outro circuito que provoque um atraso de sinal, tornando mais versátil a aplicação do AEP-FF em diferentes circuitos. O AEP-FF utiliza um número reduzido de transístores, o que resulta num baixo custo de área na sua implementação, e a sua aplicação tem pouco impacto na performance do circuito alvo, pois o sensor fica desconectado das linhas de sinal. Com o sensor de envelhecimento integrado no flip-flop é mais fácil a sua inserção nos circuitos alvo. 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.

Várias simulações em SPICE foram realizadas com tecnologia CMOS de 65nm, recorrendo a modelos "Berkeley Predictive Technology Models (PTM)", para caracterizar em detalhe o sensor e compará-lo com outras soluções já propostas. As simulações demonstram as vantagens do novo sensor face aos outros sensores de envelhecimento e a sua operação é validade em testes com circuitos de referência. É demonstrado que as degradações provocadas por envelhecimento e variações de PVT (Processo, Tensão de alimentação e Temperatura) contribuem positivamente para a capacidade de previsão de erros do sensor.

PALAVRAS-CHAVE: Sensor de envelhecimento, NBTI, Previsão de Falhas de Performance, Detecção Preditiva de Falhas.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AEP-FF	Adaptive Error-Prediction Flip-Flop
ASIC	Application Specific Integrated Circuit
BIST	Built-In Self-Test
BTI	Bias Temperature Instability
CGC	Complementary Gate Capacitor
CLK	Clock
CME	Critical Memory Element
CMOS	Complementary Metal-Oxide Semiconductor (logic family)
CP	Critical Path (higher propagation time path)
CUT	Circuit Under Test
DE	Delay Element
DE_L	Delay Element for Lower delay times
DE_M	Delay Element for Medium delay times
DE_H	Delay Element for Higher delay times
DG	Double Gate transistors
DLL	Delay-Locked Loop
DRAM	Dynamic Random-Access Memory
DVFS	Dynamic Voltage and Frequency Scaling
DVS	Dynamic Voltage Scaling
DW	Digital Word
EM	Electromigration
EMI	Electromagnetic interference
FET	Field-Effect Transistor
FF	Flip-flop
Fin-FET	Fin Field-Effect Transistor
FPGA	Field Programmable Gate Array
H	Hydrogen (chemical symbol)
HBD	Hard Breakdown

HCI	Hot Carrier Injection
HK+MG	High-K Metal Gate
IC	Integrated Circuit
IT	Interface Traps
MC	Monte-Carlo
MOS	Metal-Oxide Semiconductor
MOSFET	Metal-Oxide Semiconductor Field-Effect Transistor
NBTI	Negative Bias Temperature Instability
NC	Nominal Conditions (circuit operation)
NMOS	N-type Metal-Oxide Semiconductor (transistor or logic family)
NMOSFET	N-type Metal-Oxide Semiconductor Field-Effect Transistor
PBTI	Positive Bias Temperature Instability
PFD	Predictive Fault-Detection
PFP	Performance Failure Prediction
PI	Primary Input
PM	Pipeline Multiplier
PMOS	P-type Metal-Oxide Semiconductor (transistor or logic family)
PMOSFET	P-type Metal-Oxide Semiconductor Field-Effect Transistor
PO	Primary Output
PTM	Predictive Technology Model (SPICE transistors models)
P	Process
PV	Process and power-supply Voltage
PVT	Process, power-supply Voltage and Temperature
PVTA	Process, power-supply Voltage, Temperature and Aging
R-D	Reaction – Diffusion (model)
RTL	Register Transfer Logic
SBD	Soft Breakdown
SC	Stability Checker
Si	Silicon (chemical symbol)
SiO2	Silicon Dioxide
SOI	Silicon On Insulator

SPICE	Simulation Program with Integrated Circuit Emphasis
SRAM	Static Random-Access Memory
T	Temperature
TG-MSFF	Transmission Gate Master Slave Flip-Flop
TDDB	Time Dependent Dielectric Breakdown
UTB-SOI	Ultra-Thin-Body Silicon On Insulator
V	Power-supply Voltage
VT	Power-supply Voltage and Temperature
VTA	Power-supply Voltage, Temperature and Aging
Vth	Threshold Voltage
WCC	Worst Case Conditions (circuit operation)

1. INTRODUCTION

Current electronic circuits are mainly built with Metal Oxide Semiconductor Field Effect Transistors (MOSFET), where p-channel MOSFET (PMOS) and n-channel MOSFET (NMOS) are used globally in digital circuits. Together they form the widely used Complementary MOS logic (CMOS) and it is being used in nearly every digital electronic product worldwide. CMOS technology features low static and dynamic power consumption, high switching speeds, high density integration, high yield, low cost production and reliability, and is scalable.

In fact, scalability may be considered the characteristic that turned this technology a success and its principle states that by reducing the circuit transistors dimensions by a constant proportionality factor [1] [2], its main characteristics would change in a predictable way, directly proportional to the applied factor, and force Moore's law [3] to be kept up to date, not only in the number of transistors used, but also in performance. Over the last 10 years, digital CMOS circuits have evolved continuously to smaller circuits, being current nanometer technologies able to build circuits where transistors channel lengths are inferior to 30nm. By the time this work is being written, major semiconductor companies are currently selling their products manufactured with 22nm CMOS technology [4] and announces 10nm technology to be ready in 2015[5].

Ultimately the density scaling is the driver of CMOS scaling, and not the performance enhancement. Of course one tries to get the maximum performance at a given generation, but the productivity increase (i.e. lowering production cost) and getting more devices in the new multi-core era is an absolute requirement [6]. Over the last 15 years, there has been a new CMOS technology node approximately every two years. The key feature of every node has been 2X density shrink and ~35% performance gain per technology node [6][5].

Higher performance, lower power consumption, smaller devices and higher integration represents the main targets for continuous CMOS scaling. However the road to keep pushing technology is not straight forward and reliability issues arises every time transistors dimensions are reduced. Current new nanotechnologies represents the state of the art in CMOS digital circuits miniaturization but are also the most challenging technologies when it comes to reliability issues. Potential reliability mechanisms, identified in older generations, have higher probability to become failure causes in nanometer technologies, and new devices face higher probability of early-life failure and lower operation lifetime [7][8][9]. The reduction of the transistor's size brings additional problems, as transistor's conductivity is now being assured by fewer doping atoms, making the circuits more vulnerable to external influences. The increasing variation of a given parameter, in nanotechnologies, leads to increased uncertainty in circuit electric behavior. As a consequence, variability is becoming one of the leading causes for chip failures and delayed schedules [10], and causes increasing uncertainty in system behavior, namely on its performance. Hence, lower circuit's dependability and reliability [11] [7] are to be expected, when moving to low nanometer range.

However, special safety-critical applications, like in automotive electronics, medical equipment, or aerospace applications, system errors, especially due to performance failures, are not acceptable, as human lives are at stake [12] [13]. In these applications, the quality of a manufactured part is more difficult to assess. Hence, digital system testability in nanometer CMOS products is a serious problem, requiring delay testing [14]. Additionally, even assuming that fault-free manufactured parts are shipped to the customers and used in more complex hardware/software systems, system operation may accelerate the activity of physical defects, unscreened during production test (weak CMOS). Lifetime test may be required, especially for safety-critical products, to detect the emergence of dependability (timing) failures.

In fact, there are a number of effects that can cause errors in a circuit, permanent or not, and the most important ones are: crosstalk [15] [16], supply voltage overshoot/undershoot, reflection, electromagnetic interference (EMI) [17], power-supply noise [18], signal skew [19] [20], and aging [21]. In this work, we are concerned with the disturbances that cause the unexpected increasing of propagation delays and may lead to a delay-fault, especially if long term cumulative effects are

also considered. The most studied effects are typically Process, power–supply Voltage, Temperature and Aging (PVTA) variations. Process variations, VDD depletion, ground elevation, higher temperature[1] of operation may cause propagation delay through signal path increase (in common CMOS circuits) and, ultimately, cause a delay-fault. Moreover, aging effects caused by phenomena like Negative Bias Temperature Instability (NBTI) (the dominant long-term effect in nanometer CMOS technologies [22]), Hot Carrier Injection (HCI), or Time Dependent Dielectric Breakdown (TDDB), among others, are gaining increase relevance in new technologies and degrade circuit performance over time. All these (operation-dependent or not) variations may lead to chip failures, especially if several effects are considered simultaneously, or if cumulative degradations occur (aging effects).

To deal with this variability, designers use conservative approaches, with wider relative circuits' time slacks (error margins). However, these conservative approaches lead to non-optimized circuits. During the different stages of the design flow of a chip, it is possible to reduce the initial huge time slacks as physical implementation stabilizes some variability. Nevertheless, operation-dependent variations and aging degradation imposes the need of conservative error margins, if error must be avoided (like in safety-critical applications).

From the design point of view, well-known techniques [23] are routinely used to limit the impact of VDD variations on clock generation and distribution performance. Other known techniques can be also used to reduce the impact of PVT (Process, power-supply Voltage and Temperature) variations in the operation of digital circuits, like [24], [25] and [14]. Furthermore, various aging sensor topologies have been presented, to globally detect circuit's performance degradation [26], or to locally monitor critical path (CP) delay enhancement with aging [27]. The concept of local performance sensors can also be used to prevent harm, with a Performance Failure Prediction (PFP) methodology [28]. The underlying idea is to detect when critical paths (CP) propagation delays exceed a user-defined time threshold limit, τ_{th},

[1] Nanometer technologies have started to exhibit inverted temperature behavior [79] [80] whereby worst timing may be achieved at low voltage, slow process and low temperature. The inverted temperature effect varies for each library cell, and gets worse when lower operating supply voltages are used.

considered as dangerous in safety-critical applications. Hence, PFP may be viewed as a predictive delay fault detection technique [28] and will be addressed in this work.

1.1 PROBLEM ANALYSIS

In new node nanometer technologies, circuit aging is of great concert when talking about reliability and performance [21]. Circuit's aging is characterized as a long term gradual degradation process that may affect circuit's performance during its lifetime, and also trigger physical defects latent from manufacturing [29]. CMOS aging affects performance by slowing down circuits over time, and its effects are cumulative and cause permanent damage. The causes for circuit's aging includes BTI (Bias Temperature Instability), HCI (Hot-Carrier Injection), Electromigration (EM) and TDDB (Time Dependent Dielectric Breakdown) [30], where NBTI (Negative Bias Temperature Instability) is recognized as the main aging degradation mechanism in current nanometer technologies [31][32][33]. It affects mainly PMOS transistors and results in gradual absolute threshold voltage (|Vthp|) increase over time, decreasing PMOS switching speed [21].

Giving particular attention to synchronous digital circuits, they represent the most common type of digital circuits today. This sort of circuits use a clock signal to coordinate and synchronize all the signals traveling within the circuit, making every signal transmission between internal memory elements to occur at known, predefined, moments. This makes possible to predict and easily coordinate a complex circuit operation, characterized by multiple connections and multiple data exchanges between its internal modules. Synchronism provides predictability and simplifies communication between systems.

Unfortunately, synchronism also brings vulnerability to circuits and, if it fails, performance errors may occur, as they are particularly sensitive to variability due to their dependence in signals-clock synchronism to work properly. Synchronous digital circuits are built with memory elements (Flip-flops) interconnected by combinational logic, where data transactions between memory elements are triggered by a common

clock signal. The time required for signals to travel between two memory elements depends on the combinational logic path between them, and the slowest path defines the maximum operating frequency. Ideally, the clock period must accommodate the longest signal propagation time, in the whole circuit, plus the memory element setup-time, to assure a stable value during setup-time and so guarantee a correct data capture. If a combinational path propagation time increases due to variability, the signals may arrive within the flip-flop setup time, or after, and a delay fault occur resulting in corrupted data or circuit failure [34].

However, safety-critical products (e.g., safety and alarm systems, automotive industry, medical equipment, aeronautic industry, etc.) demands fully reliable circuits [27]. Usually this is accomplished at performance or area cost. The first solution used in synchronous circuits is to perform simulations in different design corners, to define a worst case delay deviation in the critical paths of a circuit. Then a generous time-slack[2] is used to accommodate all the predicted variations. But in nanometer technology circuits, this safety margin may not be enough, especially when considering several simultaneous effects, like PVTA variations. And, this approach does not avoid errors, as they can always occur if working conditions drift away from the predictions.

To get the most out of the technology potential, new methods have being suggested and used. One technique dynamically adjusts transistors body bias and power supply voltages over time, to accommodate variations due to process variations and aging [35]. This solution allows improved performance but do not avoid errors from happening. Besides, the circuit adjustments are not based on real-time in-circuit local measurements, but on values stored to be applied at predefined times, or on measurements on a critical path replica.

Another method uses error detection circuits, sometimes aided by additional logic to recover the system status after an error occurrence [36]. This method may provide error detection but is hard to implement recovery mechanisms in complex synchronous circuits, and again it does not avoid errors from happening.

[2]Time-slack is defined here as the last unused slack of time in a period, or the difference between the clock period and the propagation delay of the longest combinational path.

To better address aging problems researchers are looking for alternative approaches, trying to better measure and handle the aging effects, looking for solutions that could potentiate nanometer technologies performance together with its reliability. New solutions includes: the uses of distributed temperature sensors to analyze performance lost; error detection sensors plus recovery mechanisms; on-line aging monitoring sensors together with external measurement equipment to better analyze a circuits aging behavior; isolated circuits, like ring-oscillators, to measure temporal circuit degradation; aging sensors integrated in a circuit failure prediction system [31].

From all this solutions, the most young and promising are the ones based on failure prediction, to predict performance errors and activate self-healing mechanisms before errors even occur. The gradual nature of aging degradation makes possible to evaluate how the circuits' performance degradation will evolve over time and predict when a delay fault may occur. This makes possible to have circuits working near their temporal performance limits, without failures and avoiding the use of worst case safety-margins. Self-healing mechanisms may reduce circuit's operating frequency or send an alarm signal to the device user. This methodology uses a new type of sensors, named aging sensors or performance sensors, to measure and monitor time degradation on critical paths.

This concept of failure prediction was first introduced by M. Agarwal [37] and uses distributed aging sensors that will locally measure the time degradation of a critical or near-critical path. Their major application was to reduce the pessimistic worst-case delay to accommodate PVT variations, which significantly limits system performance. More recently, an aging sensor methodology focusing a different application was presented in [27][28][83][84]. Here, aging monitoring is performed during product lifetime under heterogeneous VT (power supply Voltage and Temperature) variations. The methodology includes (a) programmable aging sensor design, resilient to aging and showing low sensitivity to PVT variations, and (b) a monitoring procedure, with automatic sensor insertion.

Nevertheless, the proposed methodologies have several disadvantages that limits their application, namely: (i) the sensors' circuitry must have better performance and

be less sensitive to PVTA variations than the circuit being monitored; (ii) the local sensor solutions that use a guard-band period[3], require this time window to be synchronous with the circuit's clock and to be very stable under PVTA variations; (iii) the sensors can only be active during short time periods to avoid self-aging and degradation, resulting on blind time periods were critical paths activation are undetected; (iv) sensors connected to memory elements inputs, monitoring the critical paths outputs, will affect the path's timing characteristics, reducing circuits performance; (v) the trigger signal for the guard-band window must be routed to all the sensors on the circuit and be synchronous with the clock signal, requiring one additional clock signal to be routed during design.

However, expectations are for this sort of solutions to be implemented as standard elements in future integrated circuits.

1.2 OBJECTIVES

The main purpose of this work is to develop a new aging sensor circuit, to predict the occurrence of delay-faults. The concept behind circuit failure prediction has a great potential to become part of the solution to improve nanometer technologies, and sensors design represents an important part of this new approach.

The first objective is to design a new sensor to overcome previous sensors' limitations. The new sensor must be reliable, use aging and PVT variability on its favor, be small, with low power consumption, simple to be inserted and tuned on target circuits, easy to use by failure prediction mechanisms and circuits, capable to be always on or just active at predefined periods of time, and simple to be inserted locally at circuit flip-flops, were the synchronizations errors occur. To accomplish this, concept circuits will be developed aided by SPICE simulations to find a solution circuit schematic. At the end of this step a SPICE library, with the sensor sub-circuits, must be ready.

[3] Time period at the end of the clock cycle in which abnormal delays are observable.

The second objective is to prove the circuit's capabilities, through SPICE simulations were the sensor is inserted on test benchmark circuits and tested under PVTA variations.

The third objective is to compare the sensor sensibility detection capability with other aging sensors, to measure the differences, once more through SPICE simulations.

If the sensor characteristics analysis reflects a true innovation circuit approach, the fourth and final objective is to submit a patent request for the circuit.

1.3 CONTEXT OF THE RESEARCH WORK

The research & development work of this thesis was carried out in the Instituto Superior de Engenharia (ISE), University of Algarve (UAlg), in a strict collaboration with INESC-ID in Lisbon. The work team formed in both portuguese institutions, and working in collaboration with other foreigner R&D institutes and universities, namely University of Vigo is Spain, the INAOE institute in Mexico and PUCRS university in Brazil, has been developing in the last 4 years some research work on aging sensors, both for ASIC (Application Specific Integrated Circuit) and for emulated circuits in FPGAs (Field-Programmable Gate Array). Moreover, 3 other M.Sc. thesis were being developed in the ISE-UAlg, that are related to this one: (i) one to develop a methodology to predict the aging degradations and the delay enhancement of each cell and path; (ii) another to compute the test vectors and test automatically the delays in a circuit subject to aging; and (iii) another to define an aging and performance sensing methodology for circuits with classic BIST (Built-In Self-Test) functionality. Furthermore, one new thesis have already been initiated at the ISE-UAlg, to continue the research work in this area, to use the aging sensors developed in the present thesis, along with global sensors, to optimize dynamically the Power and/or the Frequency of operation in electronic systems; and 1 M.Sc. thesis and 1 Ph.D. thesis are to be initiated in 2012 at Instituto Superior Técnico / INESC-ID / SiliconGate (university partner / research partner / industrial partner, respectively), also in the same 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. The local sensors defined in this thesis are, therefore, of primary importance for the development of a new DVFS (Dynamic Voltage and Frequency Scaling) methodology, compatible with PVT and aging variations.

The research work developed in this dissertation has also been validated by the scientific community with 3 international conference publications ([38], [39] and [40]), 1 journal publication ([41]) and 1 Portuguese Patent pending (with title: "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" [42]).

1.4 THESIS OUTLINE

This thesis is organized as follows.

In Chapter 2, the problem of variability in CMOS nanometer technologies is addressed. The main effects that can degrade circuits' performance are resumed, namely: Process, power-supply Voltage, Temperature, and Aging variations.

Chapter 3 presents the state-of-the-art in Aging Sensors. Both existing local and global sensors for long-term aging degradations are presented, as well as their monitoring parameters, to highlight the limitations of currently available methodologies.

In Chapter 4 the Adaptive Error-Prediction Flip-flop architecture is presented, based on the Performance Failure Prediction concept, to allow the prediction of performance errors locally, in the critical memory cells, where the errors occur. The characteristics of the AEP-FF are also defined with extensive SPICE simulations.

Simulation results are described in Chapter 5, applying the aging sensor to different benchmark circuits. Results are presented to illustrate the application of the sensor predictive fault-detection capability. The impact of PVTA variations are also computed, with the aid of an aging prediction software tool and using Monte Carlo simulations. The comparison between the existing aging sensors is also presented, highlighting the AEP-FF advantages.

Finally, Chapter 6 summarizes the main conclusions of the M.Sc. work, and points out directions for further research.

2. VARIABILITY IN CMOS NANOTECHNOLOGIES

Variability is defined as any unpredictability, inconsistency, unevenness and changeability associated with a given feature or specification [10]. It is presented in every manufactured part, whatever it is, and will affect its functionality. Variability used to be an analog circuits nuisance, but semiconductors scaling into nanometer technologies turned variability into one of the leading causes for chip failures and delayed schedules in digital circuits.

Variability effects are classified as local and global. Local variability is the deviation occurring spatially within any one chip or die, and global variability is the difference in a parameter's value between nominal identical die (whether those dies are fabricated on the same wafer, on different wafers, in different lots, or in different fabrication plants). Their causes are classified as manufacturing process, operation process and operational variability (Figure 2.1).

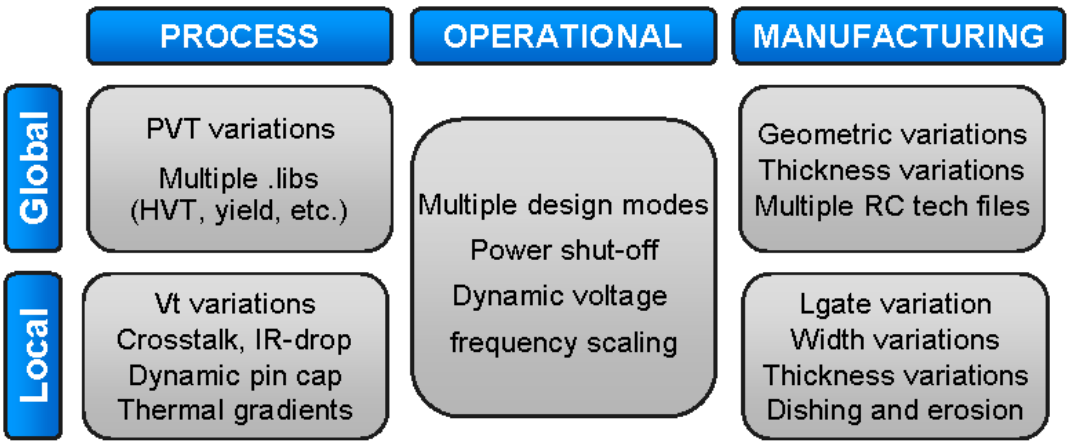


Figure 2.1. The landscape of design variability [10].

Manufacturing process variability is rooted to all the manufacturing steps limitations, resulting in transistor's oxide thickness variations, channel length and width variations, metal connections variations, among others.

Operation process variability is related to the multiple scenarios over which a circuit must operate. It will depend on environment and other external conditions, and the chip electrical signals. Ambient temperature and power supply voltage variations will affect circuits operation globally. Electrical signals circulating on each die will induce local degrading conditions like transistors threshold variation, crosstalk, IR-drop, aging, among others.

Operational variability refers to different operating modes of a chip. For example, a mobile phone chip is required to work in SLEEP, STANDBY, and ACTIVE modes. Each mode has different design constraints and the designer must assure correct operation on each mode and flawless switching between them. On very complex designs this may be difficult to achieve and unexpected system behaviors can occur, that may cause system errors.

On the current work context, operational variability will not be further explored. Regarding the other 2 classes of variability, aging (A), temperature (T), power-supply voltage (V) and manufacturing process variability (P) are considered to be the main sources of variability [2], and will be presented in more detail on the next subsections. These sources of variability are commonly referred as PVTA.

2.1 AGING

In CMOS electronic circuits, aging is the process that degrades the initial characteristics of transistors, during its life cycle. It's a long and cumulative process where the wear out of circuits occurs, caused by the time of operation and operating conditions, and triggered by several effects like temperature, power-supply voltage, humidity, operating frequency, radiation, etc. [43]. CMOS circuit aging is a gradual process that degrades circuit's performance, making it slower with time, and eventually lead to its failure. Unlike other variability causes, aging results from changes in the circuit's physical structure, which makes the degradations permanent and cumulative. In resume, aging is produced by all the mechanisms that can alter the

electric circuits' physical and electric parameters, reducing their expected life time [43].

The most important aging causes that affect nanometer technologies include [30]: Negative Bias Temperature Instability[4] (NBTI); Hot-Carrier induced degradation (HCI); Time Dependent Dielectric Breakdown (TDDB); and Electromigration (EM).

2.1.1 NBTI

Negative Bias Temperature Instability is known to be a reliability concern since 1970's [44][34], and its nature was observed since the very early days of MOS device development, in 1967 by Deal et. al., [45]. However, its effect only became significant for 130nm CMOS technologies and below [46]. Nowadays it's extremely relevant in the design of current and new analog and digital CMOS devices, as it becomes one of the major reliability concern for nanometer circuits, causing temporal reliability degradation of circuits[7][47][48][9].

NBTI is a thermally activated process that degrades PMOS transistors' physical parameters like threshold voltage and channel hole mobility [47]. It occurs under negative gate voltage (e.g., $V_{GS} = -V_{DD}$) and is measured as a gradual increase in the magnitude of threshold voltage ($|V_{th}|$) with time [9][34]. V_{th} increase results in the reduction of the absolute drain current (I_{Dsat}) and of the trans-conductance (g_m), and the increase of absolute drain off current (I_{off}) [29][34]. The reduction in drain current results in temporal degradation in the performance of a circuit causing reliability degradation over time [7].

Interface trap generation in the PMOS Si-SiO2 interface is acknowledged as the main cause for NBTI [7][49][85][47][48][34][50][9]. During manufacturing, in conventional MOSFETs, due to crystal mismatch at the Si-SiO2 interface, traps are

[4]The generic effect is the Bias Thermal Instability (BTI), which has two forms: Negative BTI (NBTI) and Positive BTI (PBTI). The NBTI effect affects primarily PMOS transistors, while PBTI affects NMOS transistors. However, for standard CMOS nanometer technologies, NBTI degradation effects are dominant when compared to other aging effects [81] [82].

present in the form of Si dangling bonds after the growth of gate oxide. Traditionally, these interface traps are passivated in ambient hydrogen to improve the device characteristics [51][52][7] (Figure 2.2).

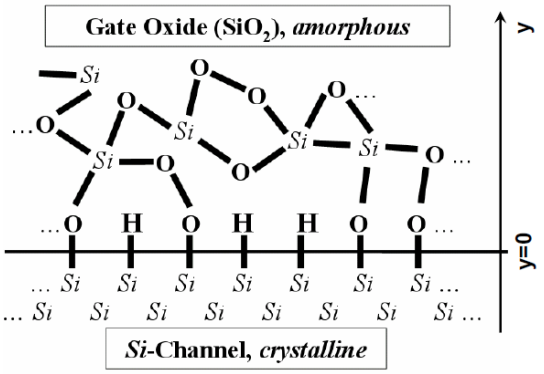


Figure 2.2.The schematic of the Si/oxide interface of a MOSFET. The dangling Si bonds are present due to the mismatch between the ordered channel and amorphous oxide. These act as interface traps, unless they are passivated by hydrogen annealing [52].

However, during transistors' operation, the Si-H bonds may be broken when the PMOS transistors are negatively biased (e.g., V_{GS}= -V_{DD}), leaving fixed positive charges (interface traps) at the Si-SiO2 interface [7][49][48][34]. The process starts with negative bias, which produces an electric field that drives inversion layer holes to the Si-SiO2 interface. The interaction of inversion layer holes with hydrogen-passivated Si atoms can break the Si-H bonds, creating an interface trap and one free H atom [7]. Atomic Hydrogen converts into molecular hydrogen (H2) and diffuses away towards the gate terminal (Figure 2.3) [47]. Diffusion of hydrogen away from the Si-SiO2 interface controls NBTI-specific interface trap generation at the Si-SiO2 interface [49].

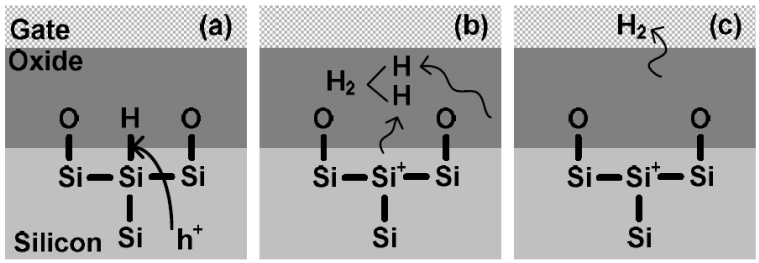


Figure 2.3. Schematic description of the reaction-diffusion model used to interpret interface-trap generation during NBTI stress [53].

Interface traps cause an increase in the threshold voltage of the PMOS transistors [48]. Either negative gate voltages or elevated temperatures can produce NBTI, but a

stronger and faster effect is produced by their combined action [34]. Typical stress temperatures lie in the 100–250 °C range with oxide electric fields typically below 6MV/cm, i.e., fields below those that lead to hot carrier degradation [34][47][50]. The trap generation process is accelerated exponentially by temperature and by the oxide electrical field [47][49]. V_{th} degradation is proportional to the number of generated traps and follows a power-law with time that is initially equal to $t^{1/4}$ and quickly evolves to $t^{1/6}$, for long stress periods [48][50]. When stress is removed (e.g., V_{GS}= 0) the process switches to a recovery phase, where some of the broken Si-H bonds are reestablished, reducing V_{th} degradation [34][48][49]. However, those hydrogen molecules that have diffused into the gate are not able to diffuse back into silicon [53]. The impact of NBTI on the PMOS transistor depends on the amount of time the device has been stressed and relaxed (Figure 2.4) [48][49]. Worst case happens with static NBTI (DC stress), where transistors are subjected to long stress periods with none or small recovery time. In dynamic NBTI (AC stress) the transistor alternate between stress and recovery phases, resulting in much less V_{th} degradation with time [49][9][21]. Reference [48] demonstrates that, for digital circuits, AC NBTI degradation does not depend on signal's frequency but on its duty-cycle.

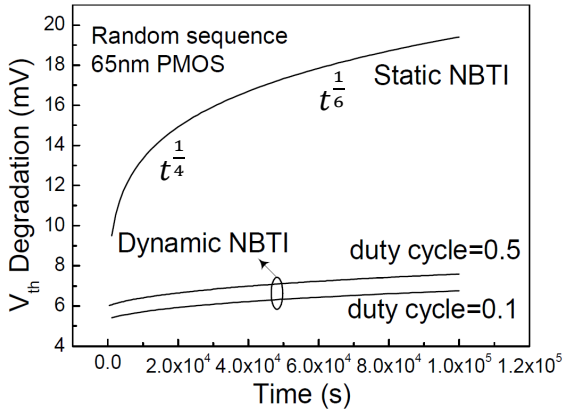


Figure 2.4. NBTI degradation over time under static and dynamic operation [21].

The power exponents reveal that the NBTI degradation is a long term process, being characterized as an aging mechanism. If dynamic NBTI is considered, then degradation times are even longer. An important detail is that NBTI induced damage, in V_{th}, is permanent and full recovery is nearly impossible.

Another NBTI consequence is that interface traps also degrades the transistor channel holes mobility, whose result is an additional shift on V_{th} [47].

NBTI became a major reliability issue in nanoscale technologies due to process changes, introduced over the last 40 years, and recent aggressive scaling of CMOS technologies to improve device and circuit performances [7][49][9]. The reduction in oxide thickness was not followed by equivalent voltage scaling, to keep oxide electric fields constant [7][49], and nanometer electric fields are higher than the reference value of 6MV/cm, as can be seen in Figure 2.5. Indeed, as gate oxide gets thinner than 4nm, the threshold voltage variations caused by NBTI has become the dominant factor to limit the transistors life time, which is much shorter than that defined by Hot-Carrier induced degradation (HCI) of the NMOS transistor [47][50]. Also, at a given oxide field, thin oxide devices were found to be more susceptible to NBTI than their thick oxide counterparts [49][34]. Thinner oxides have brought the poly-silicon gate closer to the Si-SiO2 interface, and since hydrogen diffusion through poly-silicon is faster than in oxide, scaling of gate oxides has increased NBTI susceptibilities [49].

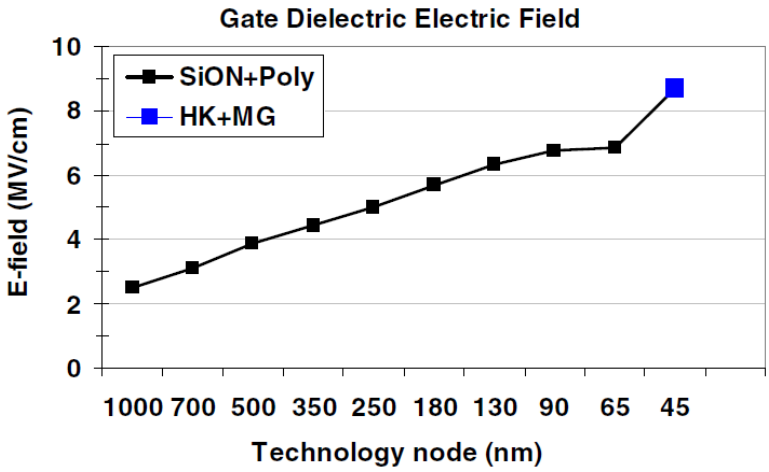


Figure 2.5. Gate dielectric electric field increase vs. technology nodes [64].

Scaled transistors are more sensitive to temperature variations [47] and the situation gets worst if we consider that variations reach up to 20ºC in the high end modern microprocessor chips [47], and up to 50ºC in DRAM, at high end microprocessors [50].

The introduction of p-poly gates for PMOS and oxide nitridation to prevent boron penetration from the gate, to reduce gate leakages, also accelerated the Si-H bond breaking at the interface [7][49][34].

Finally, the majority of digital CMOS technology requires surface channel devices to improve performance [34], which are more sensitive to NBTI than their buried channel counter parts [49].

In digital circuits, NBTI plays an important role in circuit's performance, reliability and lifetime. To understand how a circuit behaves when subjected to NBTI is a challenging task because it is almost unpredictable due to the nature of the circuit himself and NBTI characteristics. Since digital circuit consists of millions of transistors and nodes with differing signal probabilities, asymmetric levels of degradation are experienced by various timing paths [48]. Plus, the transistors on a chip die may experience different temperature levels and variations depending on their locations on chip and operation (stress/relaxation) conditions [50]. To complicate even more, NBTI shows a statistical component where circuits degrade at different paces under an identical aging scenario (Figure 2.6) [53].

Besides this, the increase in the transistors OFF current, which leads to a global increase in circuit's standby power, jeopardize its application on battery powered or energy efficient applications, a current trend in modern electronics.

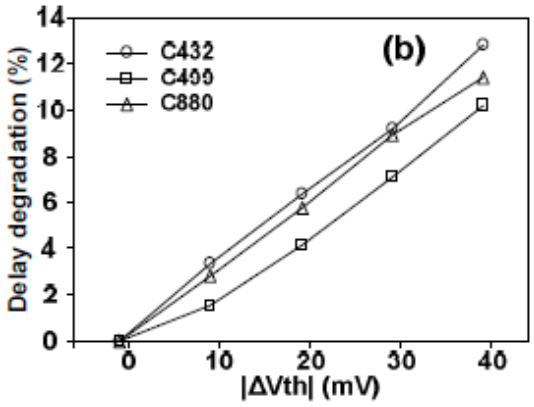


Figure 2.6. Performances of various circuits degrade at different pace under identical aging scenario [53].

The NBTI effect can be modeled according to the Reaction-Diffusion (RD) model [49][9] and its recent updates to include temperature modeling, which improved the

RD model accuracy[47][50]. With this RD model, it can be obtained a mathematical expression to calculate the drift in the V_{th} of each transistor, according to its temperature and frequency of operation, and with its workload probability. These calculations are out of the scope of this thesis, and were used in another M.Sc. thesis work [43], developed in cooperation with the present thesis work by a colleague of the ISE-UAlg R&D team, to predict the impact of aging (considering NBTI as the main degradation effect) in the circuit's path delays degradation.

Nevertheless, how much a transistor ages due to NBTI depends very much on how long it is under stress and recovery. A device under constant stress conditions, with little or none recovery time is said to be under Static NBTI, while a device whose operation switched between stress and recovery states is said to be under dynamic NBTI. In a digital circuit running a task, dynamic NBTI becomes the main degradation process, however if the circuit enter sleep or idle modes, which is common in mobile and environmental friend electronics, we may have or not a STATIC NBTI situation. Actually, it depends on how PMOS transistors status is before entering sleep or idle mode.

Static NBTI results in accelerate circuits' degradation. V_{th} change under dynamic conditions is dramatically different from that in the static mode (Figure 2.7). With the recovery in dynamic switching, ΔV_{th} due to NBTI may be reduced by 2-3 times as compared to that purely under static NBTI stress [9].

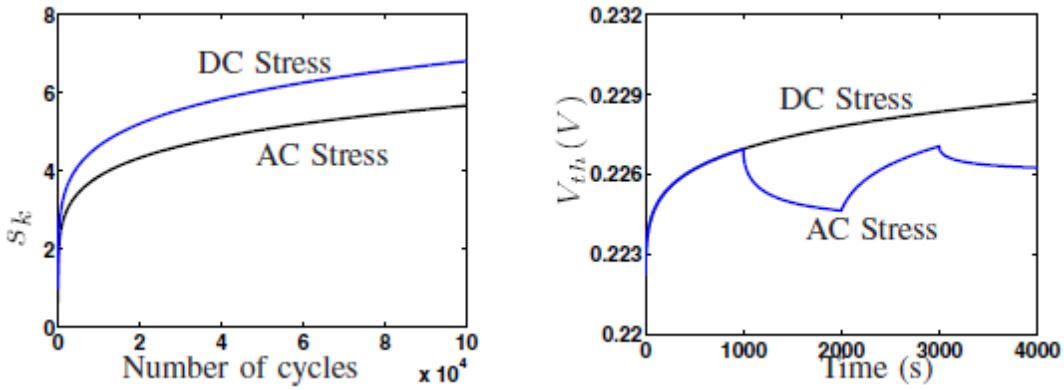


Figure 2.7. Simulation results showing NIT and VTH for DC stress and AC stress [48] .

Regarding dynamic NBTI in digital circuits, it was demonstrated that degradation is relatively insensitive to switching frequency when it is above 100Hz (Figure 2.8)

[21]. It was proven that NBTI induced degradation was frequency independence in [48], and the important parameter was the signal's duty cycle (Figure 2.9).

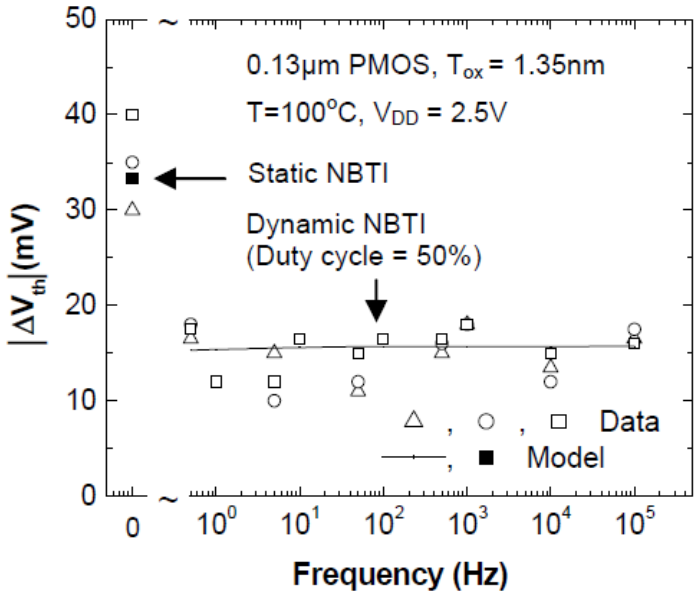


Figure 2.8.Frequency-independence of NBTI [9].

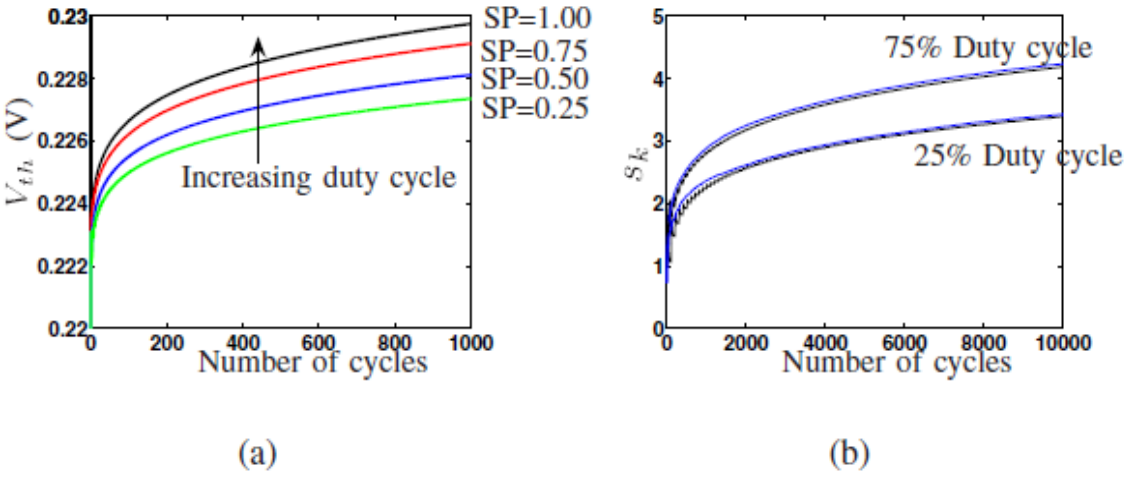


Figure 2.9. (a) signal probability dependency of trap generation shown for four waveforms of equal frequency but varying duty cycles. (b) simulations for 25% and 75% duty cycle waveforms[48].

2.1.2 OTHER AGING MECHANISMS

2.1.2.1 Hot Carrier Injection

Hot-carrier induced (HCI) degradation is a MOSFET transistors degradation mechanism where high-energy charges are injected into the gate oxide near the drain region, resulting in localized oxide charge trapping and/or interface trap generation, that gradually builds-up and permanently changes device oxide-interface charge distribution, degrading its current-voltage characteristics and threshold voltage, though its performance and reliability. HCI degradation is one of the mechanisms responsible for MOSFETs aging.

HCI became a serious problem with the introduction of submicron technologies, where the reduction of the transistors dimensions lead to increased electrical fields in the MOS channel region, both vertically and horizontally. Electrons and holes within this fields gain extra kinetic energy, high enough for some of these charges (hot carriers) to be injected into the gate oxide. Looking at the physics behind this mechanism (Figure 2.10), one of the main HCI effects is the impact ionization[5] phenomenon, which is made of two moments. For an NMOS transistor, the primary impact occurs at the drain edge, where charges arrive near the saturation velocity[6] due to the strong longitudinal electrical field in this zone. The electrons (minority carriers) are collected by the drain and the holes (majority carriers) are sent to the substrate. The holes flowing into the substrate constitute what is known as the substrate current,

[5] Impact ionization is the process in a material by which one energetic charge carrier can lose energy by the creation of other charge carriers. For example, in semiconductors, an electron (or hole) with enough kinetic energy can knock a bound electron out of its bound state (in the valence band) and promote it to a state in the conduction band, creating an electron-hole pair.

[6]Saturation velocity is the maximum velocity a charge carrier in a semiconductor, generally an electron, attains in the presence of very high electric fields. Charge carriers normally move at an average drift speed proportional to the electric field strength they experience temporally. The proportionality constant is known as mobility of the carrier, which is a material property. A good conductor would have a high mobility value for its charge carrier, which means higher velocity, and consequently higher current values for a given electric field strength. There is a limit though to this process and at some high field value, a charge carrier cannot move any faster, having reached its saturation velocity, due to mechanisms that eventually limit the movement of the carriers in the material.

and this is the main consequence of the primary impact ionization. The secondary impact ionization occurs when the holes, subjected to a high electrical field, collide with the substrate atoms, ionizing them and releasing electrons. These free electrons, subjected to the depletion zone vertical electrical field, will then drift away from the substrate towards the gate, with enough energy to be injected into the oxide, resulting in a gate current. Experimental results show that the energy produced by hot carriers, 1.5eV from the primary impact and 3eV to 3.5eV from the secondary impact ionizations, corresponds to interface state threshold energy, meaning that this effect does create interface states [54]. Interface state represents unwanted deposition of charges at the Si-SiO2 interface, which affects the expected behavior of the transistor, modifying its transconductance and threshold voltage, in a similar way to what happens due to NBTI.

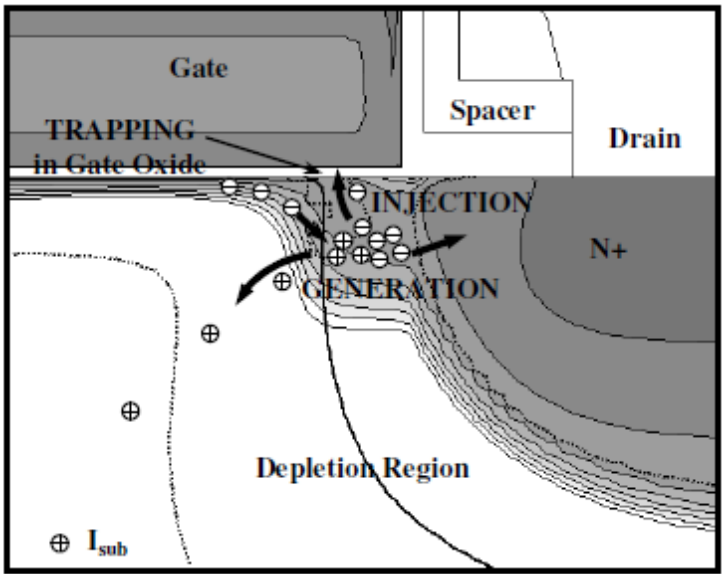


Figure 2.10. Hot carrier effects [55].

With a higher impact on NMOS transistors, this form of degradation is more pronounced near the drain region and is enhanced at saturation, when carriers gain maximum velocity.

HCI represented one of the main reliability issues in the submicron technology, but its effects are not that significant in new nanotechnologies, were the main reliability mechanism is NBTI.

2.1.2.2 Time Dependent Dielectric Breakdown

Time Dependent Dielectric Breakdown (TDDB) is a degradation mechanism characterized by failure in the MOSFET gate oxide. Due to TDDB, a conductive path forms through the oxide, between the gate and the substrate, and becomes impossible to control the current flow between drain and source [56].

The way this mechanism affects the circuits split TDDB in hard breakdown (HBD) and soft breakdown (SBD). HBD represents a catastrophic failure of the device and the entire circuit, with the formation of a good conductive path between the gate and the substrate. Soft breakdown, on the other hand, does not destroy the transistor functionality instantly but over time. Soft breakdown keeps the insulating properties of the oxide but leads to parametric variations such as energy, delay and noise margin over time. While the dielectric properties are maintained, before HBD, SBD manifests as an increase of gate current and may be modeled as a two resistors between gate and transistor's source and drain (Figure 2.11) [56].

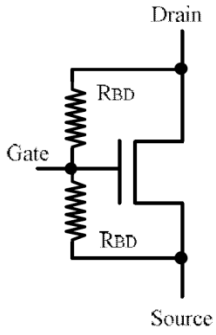


Figure 2.11. TDDB model for NMOS transistors [56].

The most widely accepted degradation model for TDDB is the "trap creation" process, where hydrogen species are freed from the silicon layers by energetic electrons and then drift and diffuse through the oxide reacting with the lattice to produce traps and other defects. When these traps and defects form a conducting across the oxide, then breakdown happens (Figure 2.12). The lifetime of a particular gate oxide is determined by the total amount of charge that flows through the gate oxide by tunneling current [56].

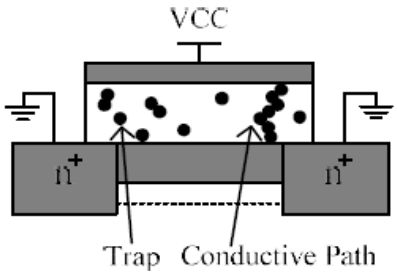


Figure 2.12. TDDB effect cross section view [56].

According to three typical TDDB models (thermochemical, anode-hole-injection, and voltage-driven), the gate oxide break down is sensitive to gate voltage, junction temperature and gate oxide area. In digital circuits, MOSFET devices are more affected by TDDB when they are ON (Figure 2.13) [56].

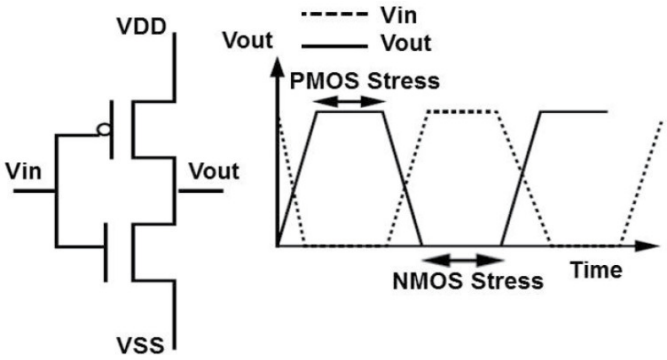


Figure 2.13. TDDB stress during steady state inverter operation [13].

Due to scaling, MOSFET devices became more prone to TDDB effects, in particular the nanoscale devices, with oxide thicknesses below 2nm.

A recent study analyses the impact of correlation between NBTI and TDDB on the performance of digital circuits, with focus on soft breakdown, acknowledging that the NBTI induced defects are somewhat identical to the defects that dominate the breakdown process. This study reveals that including this correlation in circuit analysis can lead to 9% more delay degradation than expected, with single NBTI analysis [57].

2.1.2.3 Electromigration

Electromigration (EM) is the gradual displacement of metal atoms in a semiconductor [58]. It is a reliability problem in CMOS devices, affecting circuit metal layers and vias physical structure. EM may cause circuit failure due to open-circuits or short-circuits.

"Material Migration" is a general term to englobe the various transport processes in solid bodies. Electromigration is under what is classified as "material migration caused by an electrical field". Copper or aluminum interconnects are polycrystalline materials, which means their crystalline structure is not homogeneous but granular, making them more susceptible to physical change. When current flows, moving electrons (referred as "electron wind") will interact with the metal ions in each granule boundary [59]. EM occurs when the current density in a line is sufficiently high to cause the drift of these metal ions in the direction of the electron flow. EM is characterized by the ion flux density and failures occur when there is an asymmetry in this flow. A void, created at a point where the flux of outgoing ions exceeds the incoming flux, may become large enough to open the metal line. A Hillock, created by ions piled up at a point where the incoming ion flux exceeds the outgoing flux, may become large enough to bridge adjacent or overhead metal runs (Figure 2.14)[58].

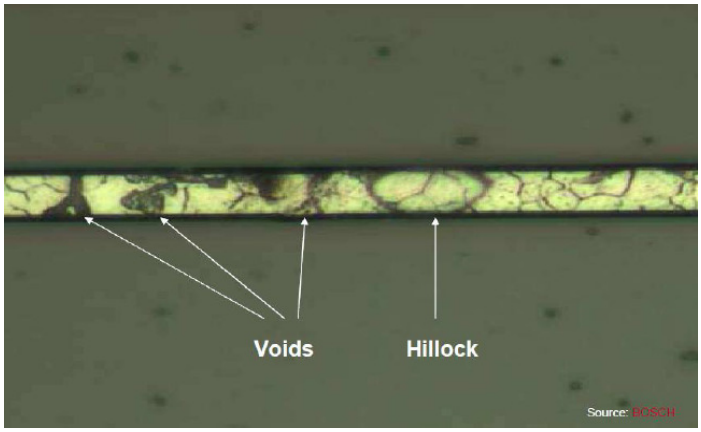


Figure 2.14. Voids and hillocks in a wire [58].

A data wire subjected to electromigration will have its resistance rising with time, until a void appears. This increase in resistance will result in higher RC time constant on the data line, leading to gradual propagation time degradation.

EM depends on temperature, current density, wire length and material. It is enhanced by higher temperature and current density. For lengths shorter than "Blech-length" (typically 10 to 100μm) EM is reduced. The wire material is also important, for example, copper has higher electromigration activation energy and higher melting point than aluminum, making copper wires capable of withstand five times more current density than aluminum wires, for the same EM degradation level [59].

High density currents expected in nanometer technologies make them very susceptible to EM induced failure. The reason for this is that dimension scaling was performed without proportional current reduction [58].

2.2 PROCESS, POWER-SUPPLY VOLTAGE AND TEMPERATURE VARIABILITY

Besides their contribution to the aging mechanisms mentioned before, manufacturing Process, power-supply Voltage and Temperature (PVT) variability also affects the circuit's performance by their own.

Process variability causes circuit's geometry and parameter variations in devices and interconnects due to photolithography proximity effects, deviation in optics, etching dependencies, polishing, and the cleaning steps. Transistors effective channel length variations, together with gate oxide film thickness variations, can result in threshold voltage and leakage current variations, affecting the circuit's performance in the same way aging mechanisms do (however, process variability is static, while aging variability changes over time). For interconnect, contact resistance variations plus metal layers width, thickness and spacing variations cause global line resistance and coupling capacitance variations, resulting in lines voltage drop, nodes capacitance and crosstalk variability [10][2]. Figure 2.15 shows how a 20x inverter chain circuit's propagation delay shifts more than 30% due to process variability. The

plot was obtained from Monte Carlo SPICE simulations (using 65nm Berkley PTM technology model), with the propagation times measured for different, random, sets of transistors L_{eff} and t_{ox}.

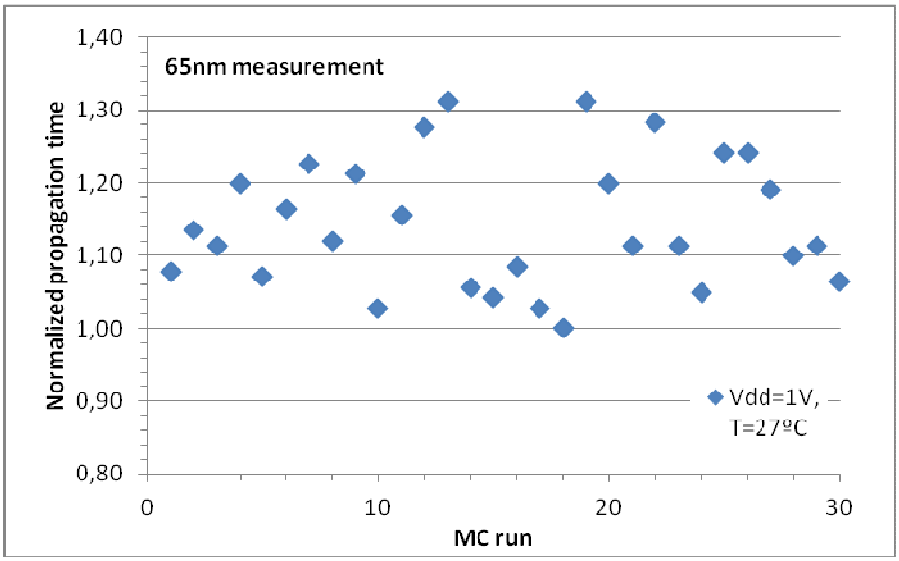


Figure 2.15. 65nm performance vs process variability.

Another example regarding process variability is presented in [2], where the operating frequency of a ring oscillator circuit shows up to 20% variations depending on their position on the wafer (see Figure 2.16).

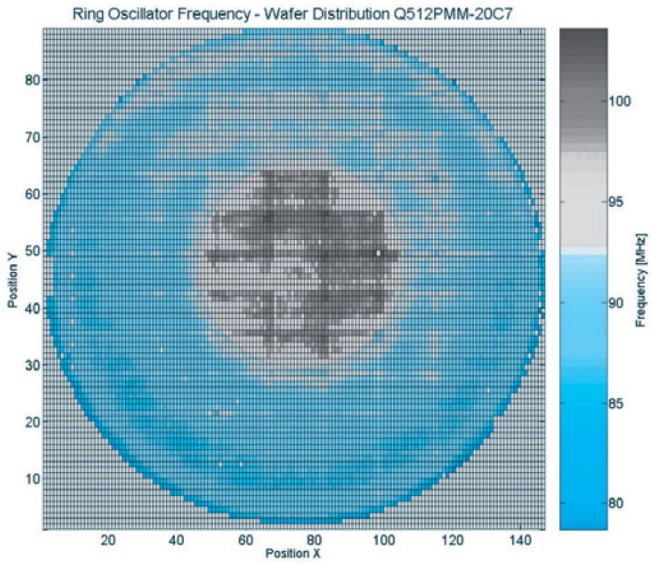


Figure 2.16. Wafer map of the frequency distribution of a ring oscillator circuit in 90-nm CMOS technology [2].

Regarding Temperature (T) and power-supply Voltage (V) variations, their impact in circuit's propagation time is visible in Figure 2.17, obtained by SPICE simulation in a 20x inverter chain test circuit, using 65nm Berkley PTM technology model. The propagation time degrades exponentially when temperature increases (20ºC to 150ºC) and supply voltages decreases (1.2V to 0.8V), up to near 200%.

Supply voltage on a circuit is not uniformly distributed and varies across the chip as well as in time. The reasons for supply voltage variations includes: the tolerance and noise of the external, or internal, power supply voltage regulator; internal IR voltage drops along the supply rails, due to the connection lines resistance; and di/dt noise.

Temperature variations in the junction, results from the sum of environment temperature and the package temperature, due to power dissipation. Because different circuit blocks on a chip have different operation activity and electrical physical characteristics, the power dissipated and the junction temperature are not uniform across the chip's die [2], resulting in different performance degradation.

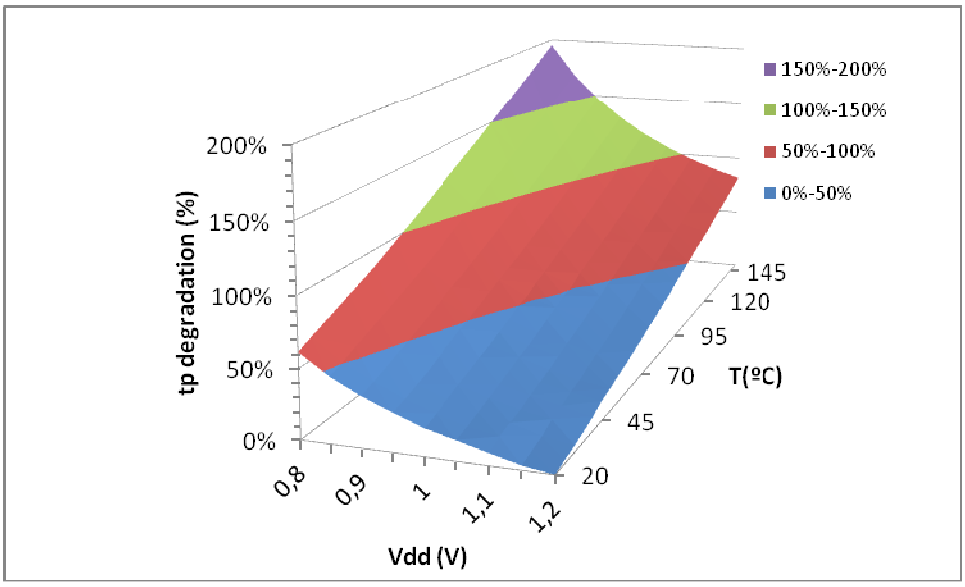


Figure 2.17. 65nm performance vs temperature and supply voltage variability

Interestingly, recent studies show the dependencies of delay degradations in CMOS digital circuits with V and T variations [14]. These delay variations can be computed at logic level, Δt_{pd}, as a function of ΔT and ΔV_{DD}, with equation (1), where $\alpha = V_{th0} / V_{DDnom}$and Ψ', η and k define charge carriers mobility variations due to V_{DD} and temperature variations, for a given technology.

$$\Delta t_{pd}(\Delta V_{DD}, \Delta T) = \frac{\psi' . \kappa(\Delta V_{DD})}{[1+\Delta T]^{\eta(\Delta V_{DD})}} \cdot \frac{\left((1+\Delta V_{DD})(1-\alpha)^2\right)}{(1+\Delta V_{DD}-\alpha)^2} \quad (1)$$

2.3 EFFECTS OF HIGH VARIABILITY IN PERFORMANCE DEGRADATION

Synchronous digital CMOS circuits represent the majority of high performance VLSI digital circuits today, like microprocessors. Synchronous digital circuits operation is controlled by a clock signal, where actions occur at exact, predefined, times, proportional to the clock period. Clock synchronization brings predictability to the circuit's operation and simplifies the circuit design and debugging of electronic systems, being one of the main reasons for the success of such typology. For these circuits, maximum operating frequency is the standard metric to quantify their performance. Higher clock signals will result in more actions per second, which is the same as higher performance. Other performance gains can be obtained using parallel processing and multi-core processors, but these are achieved by replicating the same circuit several times, and are always limited by the maximum operating clock frequency.

So, what limits a circuit's clock frequency on a digital circuit? The answer is: the slowest combinatorial logic path between two memory elements, also known as critical path (CP). Synchronous digital circuits are sequential circuits usually built with several flip-flops or latches, as temporary memory elements, interconnected with combinational logic and sharing a common clock signal. Regarding circuits using flip-flops, circuit operation is performed by transmitting data between memory elements, and with primary inputs/outputs (PI / PO). These transmissions, or flip-flops outputs' updates, occur simultaneously, at clock's rising or falling edges, and will take a finite time (t_{pd}) to propagate through the combinational logic to the subsequent memory elements. On the next clock edge, all signals at flip-flops inputs will be read and placed at their outputs, which starts a new transmission event. 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 (see Figure 2.18). Only this way it is possible to guarantee

that the flip-flop output correctly captures the input value. Additionally, the time between the clock edge and the flip-flop output update is not zero, and takes also a finite time (t_{pcq}).

Combining all this information, after a clock edge, issued to start a transmission event, the signals will require a minimum time to reach the destination flip-flop and be stable before the next clock edge, equal to $t_{pt}=t_{pcq}+t_{pd}+t_{setup}$. In a circuit, each combinatorial path will have a unique t_{pt} and, because the same clock signal is shared with all the memory elements (the usual case), the higher t_{pt_max} represents the minimum clock period (T_{C_min}) for the circuit to work properly, which is associated to the critical path. Ideal t_{pt} includes also the clock skew contribution, resulting in $T_{C\_min}=t_{pcq}+t_{pd}+t_{setup}+t_{skew}$. Figure 2.19 example illustrates this problem [2].

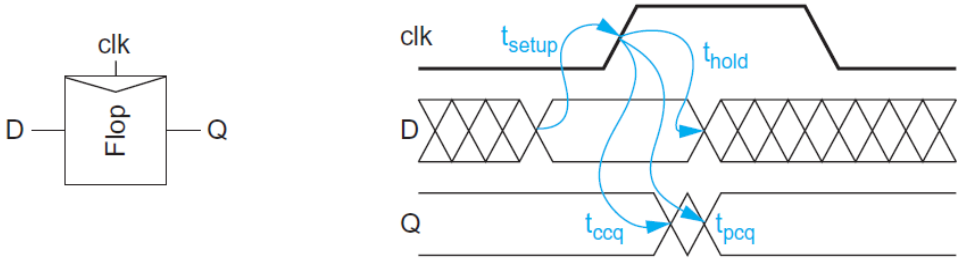


Figure 2.18. FF setup and hold time [2].

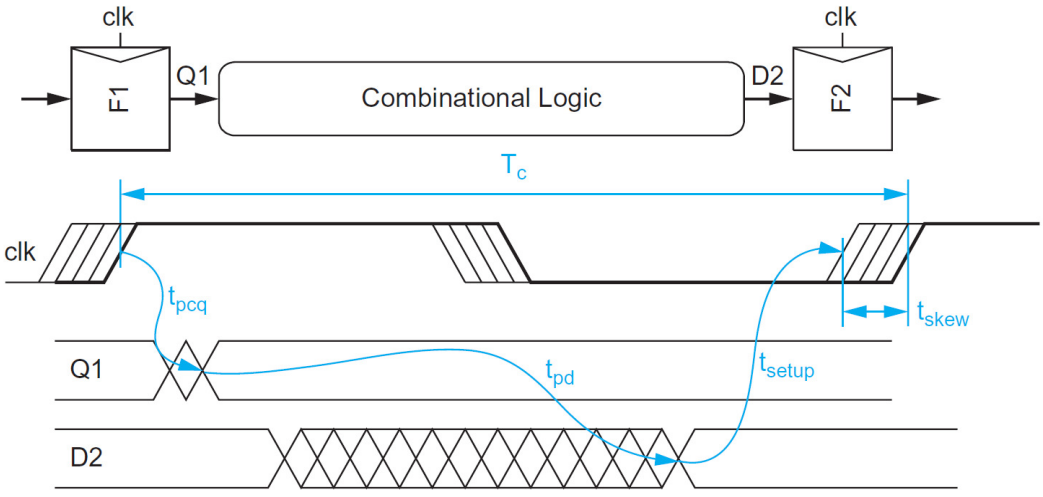


Figure 2.19. Sequential circuit temporal diagram [2].

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 FF to capture an unpredictable value and eventually cause circuit's operation malfunction and failure [33]. When this occurs, it is referred as a Delay Fault.

The vast majority of combinational circuits are composed of logic gates built with static CMOS circuits [2]. Considering the variability causes and effects, exposed in the previous subsections, a circuit path subjected to variability will suffer from additional delay [60]. 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. Finally, manufacturing process will introduce uncertainty on the nominal circuit delays between equal chips, from the same wafer or from different wafers. Aging effects may also result in circuit's paths delay rank to switch along circuit's lifetime, where a near-critical path may become critical if its total delay becomes higher than the delay of the previous CP.

Usually, the additional delays introduced by variability are analyzed at design time, and are accounted as an extra time (t_{slack}) when selecting the minimum clock period, to keep circuits reliable. This additional time slack adds performance penalties on the circuit, reducing the maximum clock frequency (equation (2)).

$$f_{clk} \leq \frac{1}{t_{pcq} + t_{pd} + t_{setup} + t_{skew} + t_{slack}} \qquad (2)$$

The amount of time slack to use will depend on the strategy adopted to deal with variability, whose topic would be explored in the next section.

During circuit design, signal path delays can be measured using Statistical Timing Analysis tools (e.g. PrimeTime™) and ranked according to their delay time. This analysis allows the designer to identify the longest paths, hence the FF where delay faults may occur, also named as Critical Memory Elements (CME). However, 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.

For a better path delay analysis, it is important that simulation tools take into account the all these variability effects. However, aging variations are difficult to predict, as they depend on circuit's workload and working conditions [40], therefore new tools and methods are needed.

3. STATE OF THE ART IN AGING SENSORS

The importance of CMOS technology scaling in the electronic devices market is reflected by the large number of researchers looking for solutions to reliability and performance problems on each new technology. The search for solutions is done at the factory, at circuit level, or even at the software level (when applicable). Process enhancements (e.g. in materials, lithography, doping methods, etching, architecture) at fabrication had allowed reliable continuous scaling since MOSFET invention. Process developments, in use or under study, includes: the use of high-k dielectrics; metal gates; modifications to the MOSFET physical structure, like silicon on insulator (SOI), ultra-thin-body SOI (UTB SOI), double gate transistors (DG), FinFET, and surrounding gate FET; the use of high-mobility material in the channel; the use of strained silicon in the channel [61][5]; solutions to reduce NBTI[62][63]. In 2007, Intel^{TM} processor scaling into previous 45nm technology was possible by replacing traditional SiO2 dielectric with polysilicon gates by High-k dielectric combined with Metal Gate electrodes (HK+MG) [64]. Current Intel^{TM} 22nm processors uses state of the art trigate MOSFET devices (FinFET with multiple fins to increase total drive strength to improve performance) [5]. Process enhancements should not be considered as a solution for any reliability issue but should be viewed as the road for evolution. A road with bumps, where the bumps represent the performance and reliability issues.

When chips production is based on a young technology node, the process is not perfectly tuned and low yield is expected. Process variability will not guarantee that an integrated circuit will be robust enough to operate properly during its lifetime. In fact, after manufacturing most chip failures occur during the first years of operation, or after a long operation time when devices start to reach their end of life. To reduce infant mortality (early-life failure), burn-in is a common solution to screen weak devices and prevent them from leaving the factory. All the chips in a wafer are subjected to high levels of stress, with high temperature and voltage, expecting for the

good devices to survive with negligible or no damage, and for weak devices to fail under such conditions [65][8].

Another traditional solution to improve circuit's reliability is to use static worst-case safety margins (slack margins). For an expected circuit's lifetime, the system's clock frequency is reduced to accommodate any shifts due to aging or operation variations and avoid delay faults. Using simulation software, with aging models, the amount of time degradation expected for the critical path is computed for a predefined period of time, with the circuit working under worst case conditions (high temperature and low voltage). This time, called the slack time, is then added to the minimum clock period resulting in an effective clock period that guarantees correct circuit operation under the simulated conditions. It is important to notice that by using generous slacks we trade performance by operation lifetime.

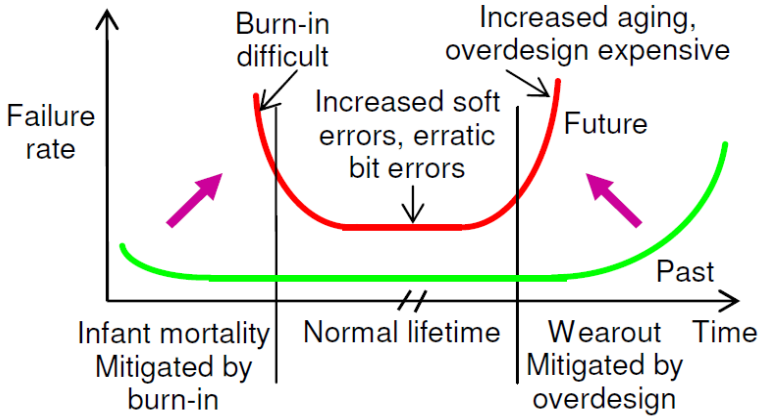


Figure 3.1. Bath-tub curve about reliability challenges in future technologies [8].

Burn-in plus worst case slack margin represent effective common practices used in pre-nanometer technologies to improve circuit's lifetime (Figure 3.1), enabling extremely low failure rates in CMOS designs [8]. However, these practices face major challenges in sub-100nm scaled CMOS technologies and are not as adequate as were before (Figure 3.1) [8]. Burn-in becomes very difficult and expensive in nanoscale technologies due to difficulties in power dissipation and high probability of circuits' thermal runaway, being possible for good devices to become damaged, suffer from accelerated aging or be destroyed [66]. This affects fabrication yields negatively and the final chip cost increases.

Inflexible designs like using worst case safety margins (generous slacks), or gate upsizing, becomes very expensive in performance loss. Considering that circuits are often designed for 11 years life span, the high variability of nanometer technologies requires elevated temporal safety margins to guarantee reliable circuit operation [31]. With this approach chips may suffer from significant reduction in speed although most of them may never be stressed to worst levels [8] [67].

These limitations turned research efforts into searching for dynamic solutions. Dynamic solutions include additional circuitry that uses parameters affected by variability as inputs to control the circuit operation on-line. This approach is further potentiated with each new technology node, where more transistors will be available per die and would become easier, and cheap, to implement extra monitoring and control circuits with low area impact.

Among all the dynamic methods, error detection is probably one of the oldest. Capable to be implemented in software and hardware, error detection identifies failures when they happen and, ideally, activate mechanisms to stop error propagation and start recovery procedures to return the system back to its normal operation. An example of error detection and recovery implementation is proposed in [36] and [68], to help reducing processor energy consumption. It is integrated in a new voltage management technique for DVS (Dynamic Voltage Scaling) processors named Razor. Razor uses delay-error tolerant flip-flops on critical paths together with in-situ error detection and correction circuits to scale the supply voltage to its limit. Error detection has a big advantage of having global applicability, where errors are uncovered independently from the error source. However, it allows errors to occur, which is a major drawback in some specific situations. Alone, error detection does not guarantee reliable operation, and is not a solution for circuits not tolerant to faults, like the ones used in the automotive, space, safety or health industries. Additionally, when error recovery mechanisms are used, the error propagation delays plus the recovery times may affect system performance significantly.

3.1 FAILURE PREDICTION

The latest and one of the most promising solutions to improve reliability and performance is the on-line circuit failure prediction [29]. Circuit failure prediction predicts the occurrence of a circuit failure before errors actually appear in system data and states [37]. The concept was first introduced by M. Agarwal as a solution to avoid worst case safety-margins, and enhance circuit's performance without losing reliability [37][67], demonstrating 4x gains in slack reduction [37]. Later, more specialized applications for this concept where proposed, like the design of ultra-reliable error-free circuits to be used in critical electronic industries [27][13], or to use it together with adaptive voltage and frequency circuits to improve microprocessors performance and energy efficiency [60]. Other examples using circuit failure prediction for performance improvement are proposed in [69], [70], [71] or [72].

The basic principle behind circuit failure prediction is to collect information about the evolution of various system parameters over time, and to analyze the collected data to predict failures [29]. For each parameter, a threshold value is defined and if this value is crossed, the system or the user will be notified of eminent failure. The threshold values will not compromise the system operation, but will be close enough to that limit. When an alarm signal is triggered, the system can take posterior actions to mitigate the degradation effects, like, for example, notify the user about the circuit's end of life, or activate self-healing mechanisms that will modify the operation conditions on the fly, for example, by decreasing the circuit clock's frequency. Data collection can be performed during circuit's operation or at predefined test periods [29].

Circuit failure prediction predicts failures caused by gradual degradation mechanisms. It cannot predict all circuit failures (e.g., radiation induced software errors or crosstalk errors), but it is quite effective preventing NBTI or early-life failure [8]. In fact, most of the publications covering failure prediction circuits are designed to fight aging phenomena, being some of these solutions classified as aging monitoring circuits, and the sensors used to collect parameter data are commonly named as "aging sensors" [29].

When compared to error detection, the major benefit arise from the fact that it enables a system to initiate corrective measures before system data and states actually get corrupted by errors, which also prevent performance loss due to error propagation and error detection latency [8] and brings to life the Performance Failure Prediction concept, explored in [13] [38].

3.1.1 MONITORING PARAMETERS

A failure prediction system's accuracy, reliability and versatility depend, first, on what parameters are chosen to be monitored. The three most used are: temperature, transistor's drain current or delay shifts [32].

The easy implementation of temperature sensors, and the fact that they are already embedded in some circuits, like high-end microprocessors, is a motivation factor to consider them as a natural source of data. In [73], a recent paper published in 2010, a method is presented to monitor the temperature in computer and network components and systems to determine their instantaneous failure rate and reliability factors, without extra hardware development or costs. However, temperature measurement alone is not a reliable metric, since aging depends on other parameters such as supply voltage or transistor's workload [29][8].

Temperature is a cause for circuit's aging but drain current variations represent an effect of such aging. Measuring one parameter that is affected by a degradation mechanism has the advantage of having all the mechanism causes embedded in it. In [74] a new sensor is proposed to track NBTI current degradation across process, temperature and activity factors. A variant of this sensor is later presented, in [32], to be used in SRAM cells. The sensor circuit is built like a 6T memory cell and the authors propose the use of this sensor to monitor SRAM memory degradation, and highlight the fact that it can also be used as a memory cell at any time, as a backup solution to replace aged memory cells. Plus, it is a small device with reduced power consumption, when compared to other solutions.

Drain current measurement is an accurate method to measure circuit aging, not being, however, an accurate metric for performance degradation measurements [53]. Performance lost results from temporal delays in data signal's propagating along a combinational path, and drain current variations on a single transistor can't represent accurately how much delay a whole path will experience, because each transistor in the path will age differently. Delay measurements in circuit combinational logic blocks is a more accurate indicator of aging [53], and is also more versatile in screening the aging process, since it will represent performance lost whatever the degradation source is.

The rest of this section will focus on delay measurement based solutions.

3.2 AGING SENSORS:

The circuits' monitoring will depend on the observation strategy. Local and/or global sensors can be used, with the imperative requirement that all are built-in the target circuit, to reduce costs and to guarantee a continuous monitoring everywhere.

3.2.1 GLOBAL SENSORS

Global sensors are used to estimate CUT (Circuit Under Test) performance degradation, due to PVTA variations. The global sensor performs a periodic delay-fault prediction, emulating CUT timing degradation. Not being possible to replicate each critical path (CP) time variation, a dummy CP is used in the global sensor. The dummy CP is specified to age faster than the circuit's CP.

Solutions for global monitoring are proposed in [75]. Here, a new on-chip sensor is proposed for AC and DC accurate NBTI aging induced degradation measurement, where the threshold voltage degradation in a PMOS transistor is translated into the control voltage of a delay-locked loop (DLL), which can be readily monitored with standard off-chip scopes, or could also be used by internal circuitry to quantify aging.

In [72], adaptive performance compensation is proposed with circuit monitoring based on global and local aging sensors, using a ring-oscillator based sensor used as a global sensor.

Kim et. al. present another aging sensor circuit in [76], whose architecture fits global aging monitoring. The sensor monitors the circuits degradation by measuring the threshold voltage difference between a NBTI/HCI stressed MOSFET and an unstressed MOSFET using an inverter chain and a phase comparator.

3.2.2 LOCAL SENSORS

Local sensors are used to perform a continuous on-line monitoring of CUT's identified critical paths, by placing local sensors at the end of each CP. Local sensors provide a fine grain aging evaluation, locally on the real functional circuit, and preventing functional errors. For continuous monitoring the sensors must be always active. The novel sensor presented in this work, the AEP-FF (Adaptive Error-Prediction Flip-flop), is a local sensor circuit. Moreover, sensors developed for local observations can be used for global monitoring, but the reverse is not possible.

To get an effective local aging monitoring, the design of local sensors turns to be very demanding. Circuit complexity, sensor output, area, power consumption, accuracy, PVTA resilience, impact on target circuit, all accounts in the final sensor's quality. Hence, sensors should be simple to insert and tune, to make them profitable for the chip manufacturer.

The sensor output will affect the control circuit and monitoring complexity. Many sensors have binary outputs for simplicity, being easier to adapt to different control circuitry. Other sensors output digitized delay measurements to provide better tracking of aging degradation.

When hundreds or thousands of sensors are needed to monitor one circuit [32], the area of each sensor is of great importance to reduce the overhead. The number of sensors to be used in a circuit depends on the monitoring strategy and sensor insertion algorithms. Local sensors are usually placed one per CME.

Sensor's power consumption should also be minimal to reduce power overhead. In fact if the whole monitoring system has a significant power consumption, global sensor temperature will tend to increase and enhance circuit's degradation.

Accuracy is necessary to avoid false predictions.

The sensors need to be resilient to process, voltage, temperature and aging induced variations, to guarantee their correct operation during the target circuit's lifetime.

Finally, direct measurements taken from signal paths should be done with little or no impact in the paths signal propagation performance.

Aging sensor circuits capable of performing local aging monitoring, based on path delay monitoring, are proposed in: [37], [53], [27], [33], [67] and [70]. All of them are designed to be connected to critical memory elements inputs and monitor the arrival times of data signals from the paths connected to those flip-flops.

The sensor architecture in [37], detects critical delays when delayed data signals arrive within a predefined guard-band time, before the clock edge that will update the flip-flop. The guard-band time is set to avoid delay faults but lower than a pessimistic WCC (Worst Case Conditions) guard-band. The sensor is composed by a delay element (DE), to define the guard-band time; a stability-checker (SC), to identify late transitions within the guard-band time; and an output latch to retain the sensor after detection.

In [27], the authors improved the Agarwal sensor's DE, to be programmable and PVTA resilient.

In [67], a smaller and faster SC is proposed to replace Agarwal's SC circuit, to reduce the aging sensor's overhead and provide faster response time when detecting data signal transitions.

The aging sensor in [33] uses Agarwal concept to assess circuit failure prediction, with a totally new circuit considered to have lower area overhead, and lower or equivalent power consumption. Moreover, this sensor includes self-checking capability to prevent possible internal faults on the sensor operation.

The sensor in [53] is another complete new circuit, with lower area overhead, lower power overhead and higher accuracy against process variations, when compared to the sensor presented in [37].

The sensor in [70] has a different approach to aging monitoring, by performing high-resolution measurements to the critical path delay. The mechanism use on-chip aging sensor embedded in CME to capture transitions and generate pulses whose width represents the actual path delay. The modulated pulses are then delivered to a control unit to accurately measure the delay.

The sensors proposed by M. Agarwal et. al. and [37] and by J. Vasquez et.al [27], will be described in detail on the next subsections. These sensors were the main references for the AEP-FF development, due to their operation principles and characteristics. AEP-FF development motivation was to find a new solution that would perform better than these two sensors and overcome their limitations.

3.3 M. AGARWAL ET AL. AGING SENSOR

M. Agarwal et al. were the first to introduce circuit failure prediction concept in digital CMOS circuit design. They introduced the concept and an application methodology to reduce circuit worst case safety margins, and proposed a new aging sensor architecture [37].

The sensor, depicted in Figure 3.2, is a modified standard latch or flip-flop with a "monitoring" circuit block which detects any 'significant' shifts in delay of the combinational logic whose output drives that latch or flip-flop. The sensor defines a guard-band time interval, before a clock rising edge, where no signal transitions should occur. If this happens the sensor's output is turned on, to signal a guard-band violation (Figure 3.3). This will not lead to a failure, because the guard-band interval is defined to be higher than the flip-flop's setup time, and the flip-flop will still continue to capture correct logic values; but it will inform the system that the combinational logic input stimulus exercises one or more paths that have aged enough to creep into the guard-band interval and is now very close from creating a delay fault.

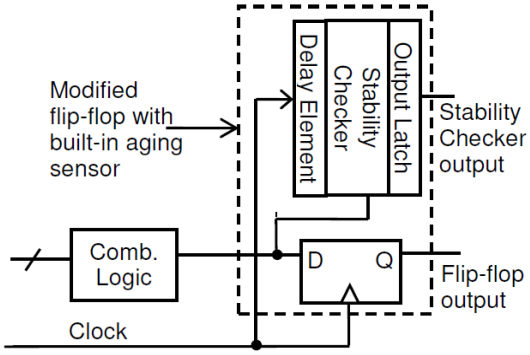


Figure 3.2. Flip-flop with built-in aging sensor.

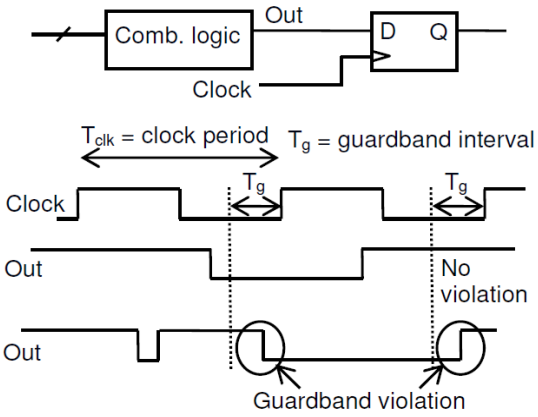


Figure 3.3. Guard-band violation due to transistor aging.

3.3.1 CIRCUIT

The aging sensor block is composed by three components: a delay element (DE), a stability checker (SC) and an output latch. The DE defines the guard-band timing, the SC will detect any guard-band violation, and the output latch will retain the sensor's output high after a positive detection.

Before explaining in detail each circuit module, notice that an important requirement for this circuit to operate correctly is to avoid aging. To make the aging sensor age resilient, the slow nature of NBTI is very valuable. It allows for the aging sensors, in particular the DE, to be turned off most of the time and avoid significant aging. During off periods, care is taken to forward bias the PMOS transistors used to define guard-band timing and detection to be off, for them not to age.

The DE circuit, depicted in Figure 3.4, is designed to be age resilient against NBTI. Monitor input will turn the DE circuit on and off. When the DE is on (Monitor=1) a negated and delayed version of the clock signal is in the output. Gate G1 will negate the clock signal (clock_b) and the total delay is set by the NOR gate G1, T10 to T13 inverter, and G2, G3 and G4 NAND gates propagation times. Delay tuning is accomplished adjusting G2, G3 and G4 propagation times, and contributions from the other elements should be negligible. When the DE is off (Monitor=0) the output is constant and set to logic level 1. All the PMOS transistors directly connected to the monitor input will age but will not degrade the delay time. This includes T14, the PMOS inside the top inverter gate and the three PMOS inside G2, G3 and G4. To prevent aging of the clock signal delay path transistors, the first PMOS transistor connected to the clock input is inside a NOR gate. I assume that the NOR is assembled like the circuit in Figure 3.5 to assure that both PMOS, M1 and M2, are off when DE is off. Next come T10, T11 and T12. The author uses T10 and T12 to isolate T11 when the DE is off, assuming that T11 will not age under that conditions. However, it all depends in the voltage levels in T11 drain and source nodes parasitic capacitances. Initial nodes levels will be set to Vdd because T11 is used as a pass transistor, enabling dynamic NBTI induced aging in T11 when the DE is off. However this should not be a problem, because leakage currents will discharge this nodes voltage removing the stress condition from T11. The last delay gates are G2, G3 and G4. The PMOS transistors in these three NAND gates connected to the clock signal will be immune to NBTI because all their inputs are set to logic level 1.

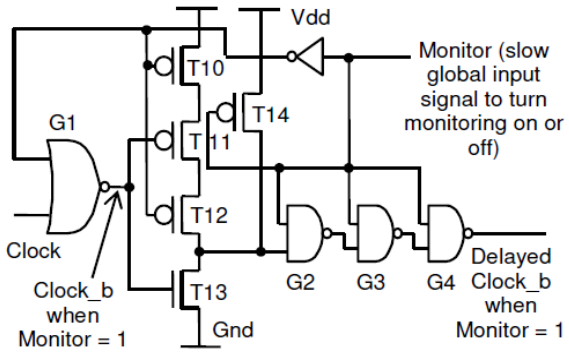


Figure 3.4. Aging resistant delay element design.

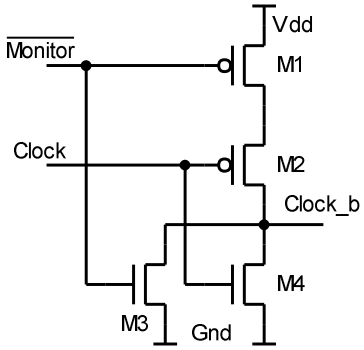


Figure 3.5. NOR gate G1.

Delay monitoring and guard-band violation detection is performed by the SC circuit (Figure 3.6). To understand its operation the waveforms in Figure 3.7 are very helpful. When clock_b = 0 (clock = 1), transistors T1 and T5 are on and both NOR gate inputs are at logic level 1, turning the SC output Out = 0. This is called the precharge phase. When clock_b turns high (clock = 0), T3 and T7 turn on, but T4 and T8 are off because the DE output is kept low for a time equal to "$T_{clk}/2\text{-}T_g$". Up to this moment transitions are allowed at the combinational logic output. When the DE output goes high starts the guard-band interval. During this guard-band T3, T4, T7 and T8 are on, T1 and T5 are off and one of the NOR gate inputs will be at 1 and the other at 0. Any transitions at the combinational logic output, during the guard-band interval, will turn the SC output high, i.e. the guard-band is violated.

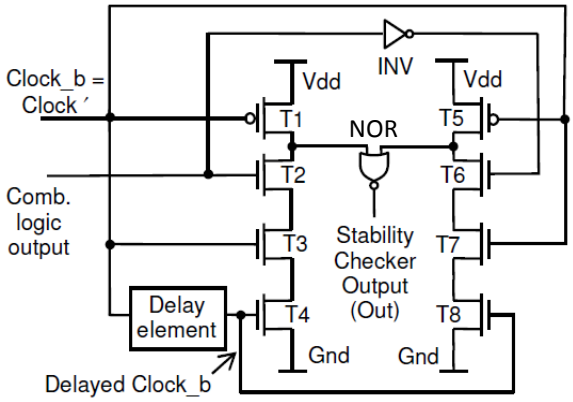


Figure 3.6. Aging resilient SC design.

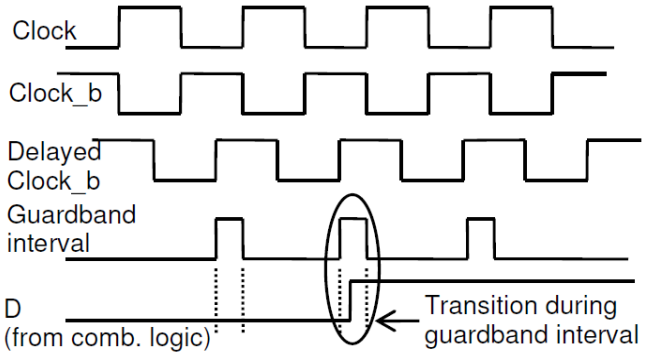


Figure 3.7. Timing diagram for aging sensor.

3.3.2 CHARACTERISTICS

To test and demonstrate the sensor characteristics, M. Agarwal et. al. performed several simulations using 65nm and 90nm technology models.

3.3.2.1 Perfomance, Power and Area Impact

An inverter chain circuit with the modified flip-flop design with the built-in aging sensor was used as test circuit to quantify the new circuit performance, power and area impact. Performance penalty represent how much the critical path propagation time will be increased with the aging-sensor electronics inclusion. Power penalty represents the power increase due to added aging sensor operation. Area impact represents the amount of extra area needed for the monitoring circuit. The simulations considered two scenarios, (i) where the one DE was used per flip-flop and (ii) where the single DE output is shared among 4 flip-flops sensors. The results returned negligible performance penalties for both (i) and (ii), lower than 1%. Power penalty is negligible if we consider that the sensor may be off most of the time, and turned on only from time to time. The results returned 0.3% of power penalty when the aging sensors are off and 12.5% when on, for (i). In (ii), the percentages reduce to 0.1% and 7.5% respectively. The area penalty depends generally on the total number

of transistor required for the aging sensors, vs. the circuit complexity. For (i) and (ii), 51 and 24 transistors are required per sensor.

3.3.2.2 Charge Sharing, Leakage and Aging

The circuits SC checker is a dynamic circuit and due to this, its reliability may be compromised by charge sharing or leakage. Charge sharing between high-impedance nodes may modify the nodes state, and transistors leakage may discharge/charge these nodes with time, if no refresh is performed on time. For this circuit no charge sharing problems exist, and no node level flip occurred due to leakage for operation frequencies as low as 2MHz. The last result is for the 90nm technology because the 65nm technology model was not available at that time. M. Agarwal et. al. expected increased leakage on future technology nodes.

Regarding aging simulation, the results returned small aging along the clock path in the DE, with 65nm technology, due to long periods in off time.

3.3.2.3 Effective Guard-band

If the combinational logic output transitions very close to either rising or falling edge of the guard-band interval, the signal transition may not be reliably detected by the stability checker (the effective guard-band is smaller than the programmed guard-band). To accommodate this unpredictability, the author found that adding 20ps to the nominal guard-band would be enough to provide reliable detections. The result is an extended guard-band time, to allow a higher effective guard-band.

3.3.2.4 Impact of process variation

The guard-band interval dependence on process variation was analyzed performing guard-band measurements in 250 Monte-Carlo simulations, with 3σ equal to 30% variation from the nominal value, in transistor channel length and threshold voltage, using 90nm technology. Results show a significant variation (Figure 3.8), up 83% around the nominal value (120ps).

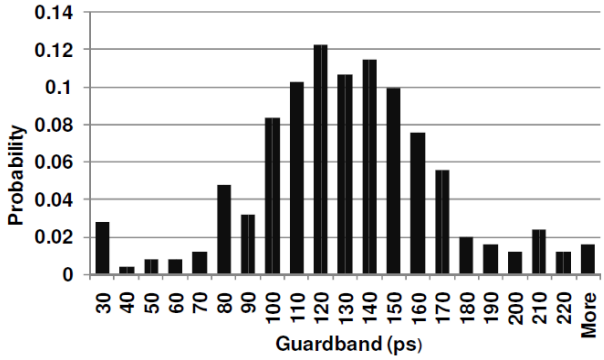


Figure 3.8. Agarwal sensor guard-band dependence on process variation.

3.4 VAZQUEZ ET. AL. AGING SENSOR

In [27], Vazquez et. al. highlight the use of circuit failure prediction to guaranteed reliable and dependable circuit operation on safety-critical systems, such as in automotive. They propose a new on-chip, on-line aging sensor to detect abnormal delays on critical paths, regardless of their origin.

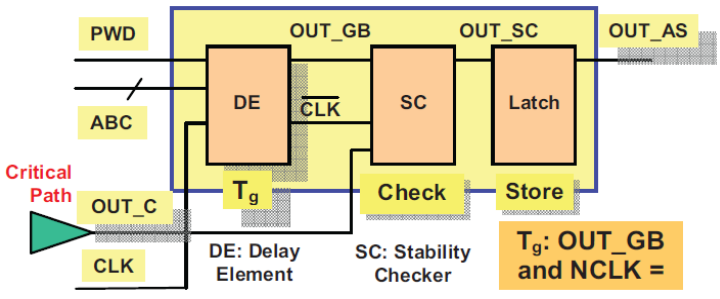


Figure 3.9. Vazquez et. al. aging sensor architecture.

The sensor has the same architecture and working principle of M. Agarwal et. al. sensor [37], but with a different and enhanced delay element circuit (Figure 3.9). The sensor is designed to have programmable guard-band time and be resilient to PVA variations. The guard-band is programmable with up to seven different values ("ABC" inputs), making possible to identify different degradation levels. Transistors dimensions avoid Lmin and Wmin to reduce the circuit sensitivity to process variations. The guard-band time depends on a Complementary Gate Capacitor (CGC) structure, to lower its sensitivity to PV variations. Aging effects are reduced by performing monitoring at predefined intervals, leaving the sensor in the off state most of the time.

3.4.1 CIRCUIT

The proposed sensor is composed of three components: one DE, one SC and one output latch (Figure 3.9). The SC and the output latch circuits remain the same as the ones used by M. Agarwal, while the DE is a totally different circuit.

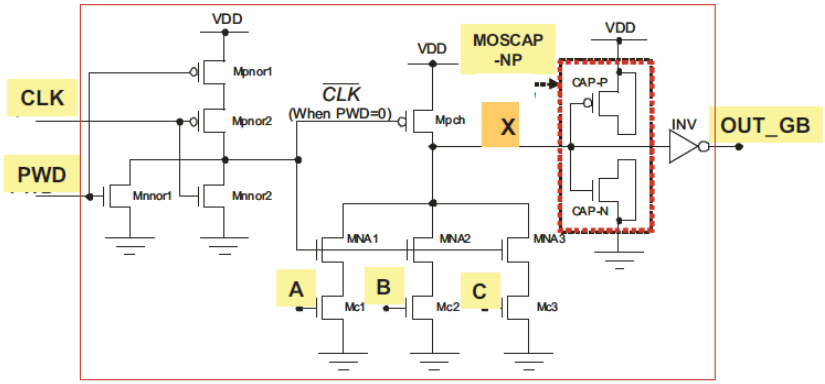


Figure 3.10. Vazquez et. al. DE circuit.

The DE circuit is depicted in Figure 3.10. The circuit uses one capacitor charge/discharge time to define the delay time, hence the sensor guard-band. The capacitor is implemented with two NMOS and PMOS transistors (CAP-N and CAP-P) in a CGC configuration referred as MOSCAP-NP. This non-linear structure was selected due to its low sensitivity to V_{DD} and process variations. To control the DE operation, PWD input is used to turn the DE on and off. When active (PWD = 0), the clock signal (CLK) is inverted by the NOR gate, resulting in signal CLKN. During

the positive clock period (CLKN = 0), Mpch turns ON and node X will be charged to Vdd (CAP-N charged to V_{DD}) and DE output goes low (OUT_GB = 0). This is referred as pre-charge phase. When CLK goes low (CLKN = 1), node X will discharge through a pull down network composed of MNA1, MNA2 MNA3, Mc1, Mc2 and Mc3(CAP-P will charge to V_{DD}). ABC inputs are combined as a 3 bits digital word, DW, and will define what transistors are on in the pull-down network. The DE delay time is related to the X node discharge time, hence programmable by DW. Programmable delay time will then result in programmable guard-band time. The use of NMOS transistors in the pull-down network makes the guard-band time resilient to NBTI effects. The waveforms in Figure 3.11 exemplify the DE operation.

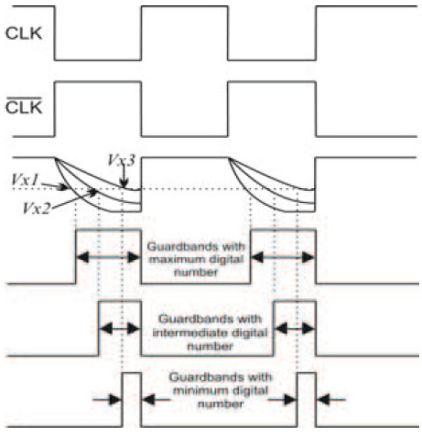


Figure 3.11. Vazquez et. al. sensor timing diagrams.

When the DE is OFF (PWD = 1), Mpch will age but CAP-P will not, guaranteeing the guard-band time resilience to NBTI effects.

3.4.2 CHARACTERISTICS

To characterize the sensor a set of simulations were performed using ST 65nm CMOS technology model, testing the circuit's functionality in a test circuit made of

cascaded identical combinational cones used in industry design (XTRAN from TecMic™).

3.4.2.1 VT variations dependence

Just like Agarwall et. al. sensor, this sensor's guard-band interval is also negatively affected by VT variations (Figure 3.12), decreasing with temperature raise and power supply voltage drop.

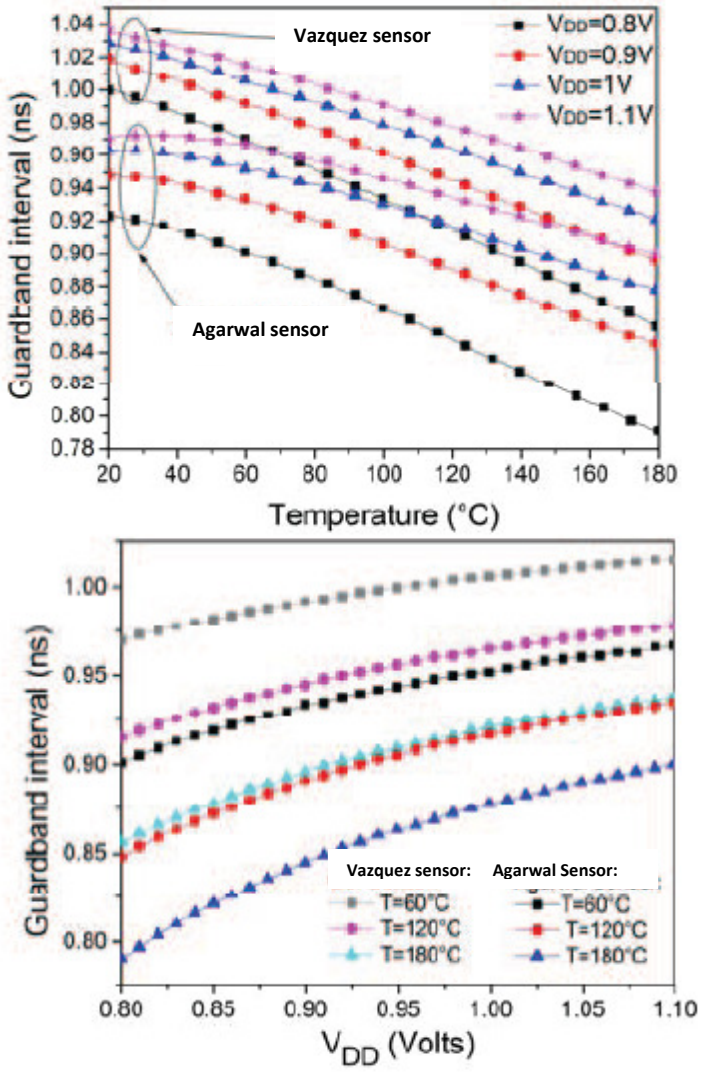


Figure 3.12. Guard-band variations with T and Vdd, for Agarwal and Vazquez aging sensors.

3.4.2.2 Programmable guard-band

For each DW word the detection range was analyzed with the sensor inserted in the circuit under test (CUT). The results, plotted in Figure 3.13, show 7 different time degradation values that trigger the sensor output, ranging from 17% (DW=7) to 42% (DW=1).

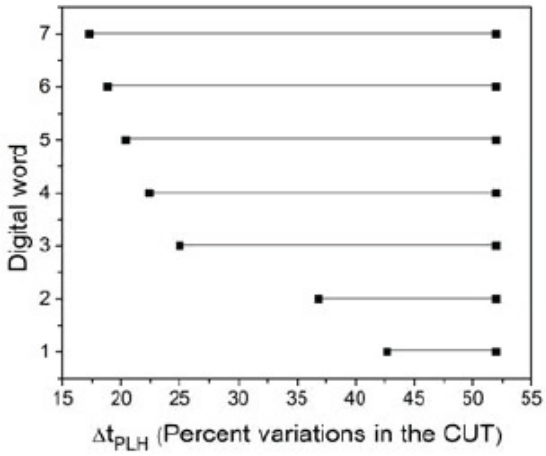


Figure 3.13. Vazquez et. al. sensor detection ranges in percent variation of the path propagation time of the CUT, under WCC (VDD=0.8V, T=180ºC).

3.4.2.3 Process variation dependence

Monte-Carlo simulations were run varying MOSFET parameters in the CUT and in the sensor's transistors to analyze the sensor operation under process variations, in WCC. Gaussian distribution with +-3σ variation of +-10% of the nominal values was assumed for three MOSFET parameters: channel length, L, oxide thickness and threshold voltage.

In Figure 3.14, MC simulations for ΔV_{thP}=15% in the CUT (Δt_{PLH}=17%), with DW=7, show effective guard-band time, considering process variations, as the minimum time between the OUT_GB late rising edge and CLKN early falling edge, when sensor output goes high (OUT_CL = 1).

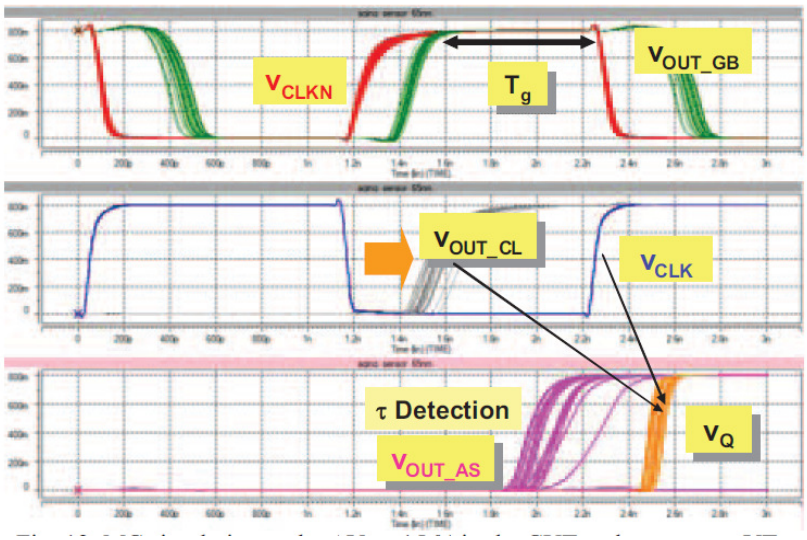


Figure 3.14. Vazquez et. al. sensor, MC simulation under ΔV_{thp}=15% in the CUT and WCC (V_{DD}=0.8V, T=180ºC)

Due to process variations, abnormal delay detections will not occur for all MC runs. The Detection Probability (DP), representing the percentage of the 30 MC runs that detect the abnormal delay, is show in Figure 3.15 for each DW word.

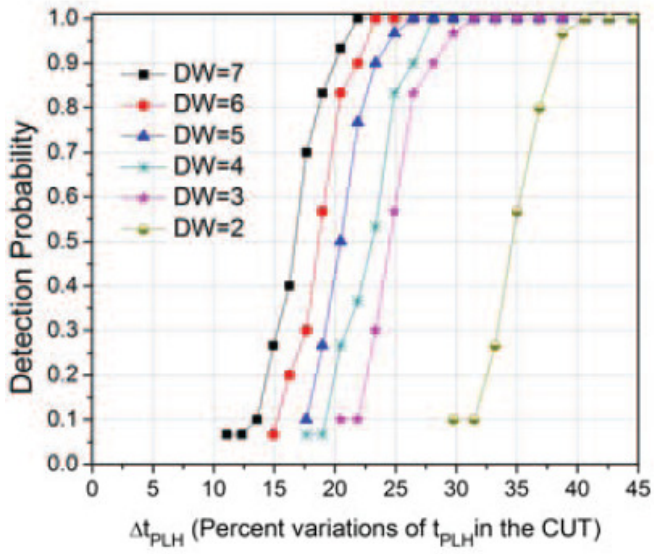


Figure 3.15. Vazquez et. al. sensor, dependence of DP on Δt_{pLH} foe each DW, under WC conditions.

4. ADAPTIVE ERROR-PREDICTION FLIP-FLOP

The present chapter presents the main characteristics of the AEP-FF, the aging sensor developed in this thesis work.

4.1 CONCEPT

The purpose of this new sensor is to monitor late transients in the data input of FF that terminate critical and near-critical paths. In previous works, like [27], a guard-band time is defined at the end of the clock cycle, defining when FF data input transients are signalized as an error prediction (late transients). In this work, there is no signal explicitly defining a guard-band time. However, a virtual guard-band time, τ_{GB}, exists, defined by the propagation delay time of a delay element (DE), which is placed inside the FF and will delay the signal data at the end of the critical path (see Figure 4.1). PVTA variations induce increased delays in the CUT data signal, increasing τ_0, and in the DE, increasing τ_{GB} (see Figure 4.2(a)). At design, a time slack, τ_{0slack}, is defined to absorb τ_0 PVTA-induced time variations. Under normal operation conditions $\tau_0 + \tau_{GB}$ will be smaller than T_{CLK}. When a worst-case operation condition (WCC) produces a timing degradation such that $\tau_0 + \tau_{GB}$ exceeds the clock period (T_{CLK}), entering the next clock positive semi-cycle, an abnormal propagation delay time is spotted as an unsafe delay, although not large enough to induce malfunction (see Figure 4.2(b)). Error prediction is performed during the clock positive semi-cycle.

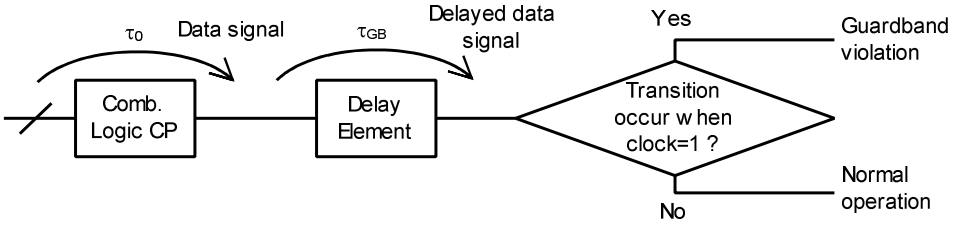


Figure 4.1. AEP-FF concept block diagram.

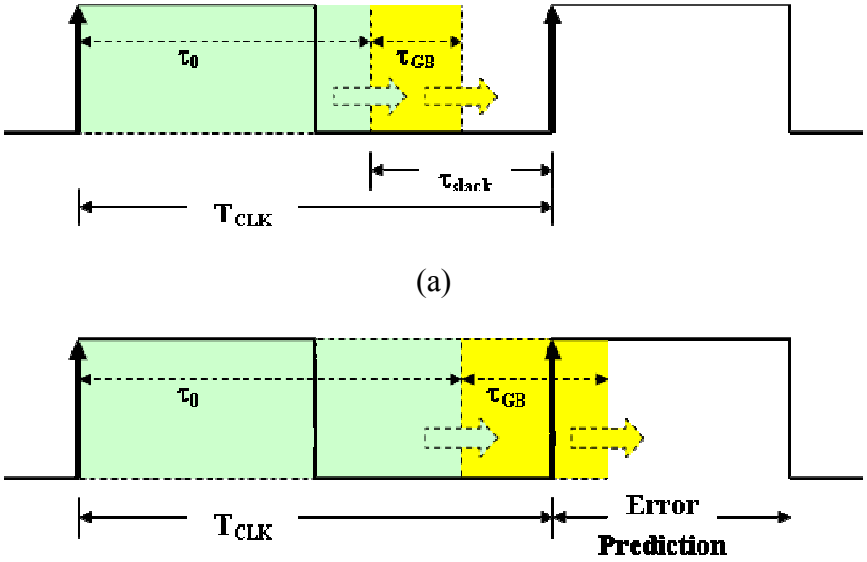


Figure 4.2. Error-prediction and sensor operation. (a) Nominal PVTA conditions, with no error predicted; (b) PVTA WCC and error prediction.

The topology of the proposed Adaptive Error-Prediction Flip-Flop (AEP-FF) is shown in Figure 4.3. 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, τ_{GB}, used by the sensor. Late transitions at FF data input (propagated to the Master Latch output) will be identified by the SC. As it will be shown later, SC has on-state retention logic, to discard the use of an additional latch to store the aging sensor output signal (AS_OUT).

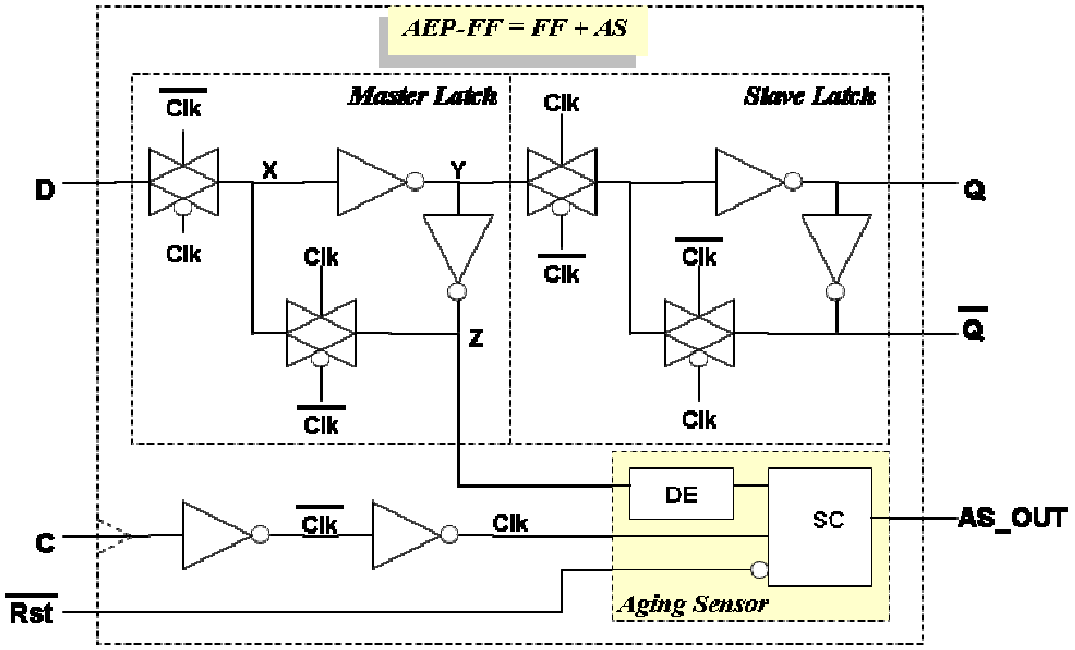


Figure 4.3. Adaptive Error-Prediction Flip-Flop topology.

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 [27][37], simplifies DE design. Basically, the new delay element is a simple buffer that introduces a delay to create a virtual guard-band where late transitions at FF data input are signalized. Moreover, if PVTA variations occur in DE, this virtual guard-band will increase accordingly. It is possible to create a power-on state for the DE, and activate it in short periods, restricting power consumption and aging effects of the aging sensor. However, even when the sensing operation is always ON, the increased workload of the DE will cause sensor´s guard-band to increase as aging effects cumulatively degrade DE performance. This way, the sensor's sensitivity is adapted with the cumulative aging degradation of the circuit. Another advantage is that the guard-band signal does not need to be distributed as a second balanced clock to the sensing FFs, as in [27][37].

The loading effect of the sensor is inside the FF; hence, it does not explicitly impact the signal path. In fact, experiments were carried out to determine the best location, inside the master latch, to connect the aging sensor (AS) circuitry. The best solution was connecting the AS circuitry at node Z (see Figure 4.3), when compared

with connections at nodes X or Y. The setup and hold times measured[7], $t_{SU} + t_H$, for connections at nodes X, Y and Z were, respectively, 55ps, 53ps and 52ps.

4.2 AEP-FF CIRCUIT MODULES

4.2.1 FLIP-FLOP

The sensor flip-flop is classified as a Transmission Gate Master-Slave Flip-Flop (TG-MSFF) [77]. When the clock signal is low, the Master Latch output Y is tracking the input D, and the Slave Latch keeps its output (Q) constant, ignoring any changes in node Y. When clock switch to level high, the Master Latch will keep its output (Y) constant, ignoring any changes in node D, and the Slave Latch will set the output Q with the value in node Y.

4.2.2 DELAY ELEMENT

Three circuit architectures are proposed for the DE module, DE_L, DE_M and DE_H, designed to use the minimum number of transistors and provide significant delay time differences between them (from Figure 4.4 to Figure 4.6 the delay time increases). The DE architecture should be chosen according to following factors: the clock frequency, the Tslack/TCLK ratio, the technology, and the sensor's sensitivity (or the PVTA WCC where the sensor starts to flag a late transition). As an example, considering τ_{slack}/T_{CLK}=30% and a 65nm Berkeley PTM technology, typically architecture (a) can be used for frequencies above 1GHz, (b) from 400MHz to 1GHz, and (c) bellow 400MHz. Moreover, as changing W/L transistors ratios also change the sensor's effective guard-band, τ_{GB}, the DE can be optimized by design. Note that,

[7] Values measured in HSpice simulations.

unlike in [27], τ_{GB} is not programmable, it is defined at design time. However, τ_{GB} is adaptive with PVTA variations, enhancing sensor's detection sensitivity. We refer to it as a *virtual guard-band* because there is no signal explicitly representing the observation interval. Each sensing FF will have its own unique PVTA-dependent guard-band (each local DE may age differently).

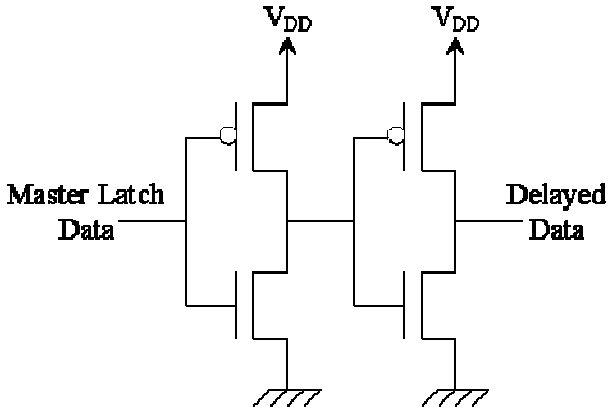


Figure 4.4. Delay element typical architecture: Low delay - DE_L.

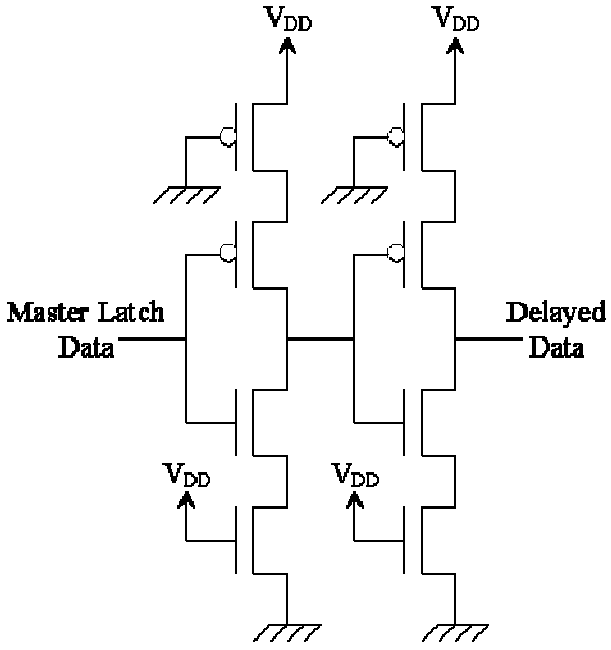


Figure 4.5. Delay element typical architecture: Medium delay - DE_M.

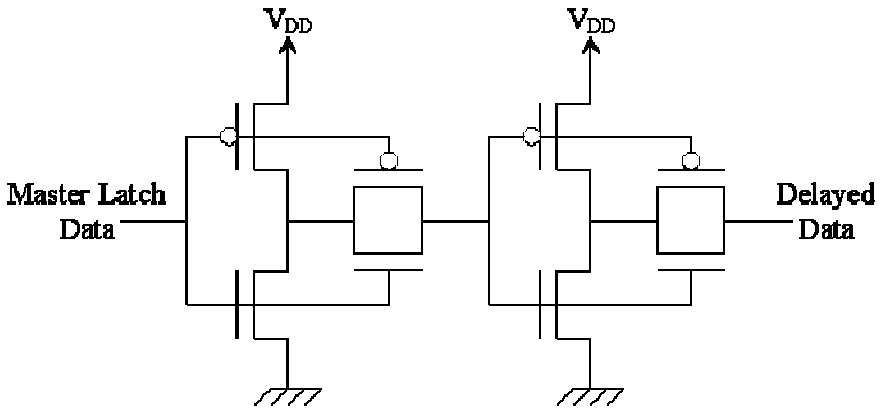


Figure 4.6. Delay element typical architecture: High delay - DE_H.

4.2.3 STABILITY CHECKER

The novel Stability Checker (Figure 4.7) is implemented with dynamic CMOS logic and has built-in on-retention logic.

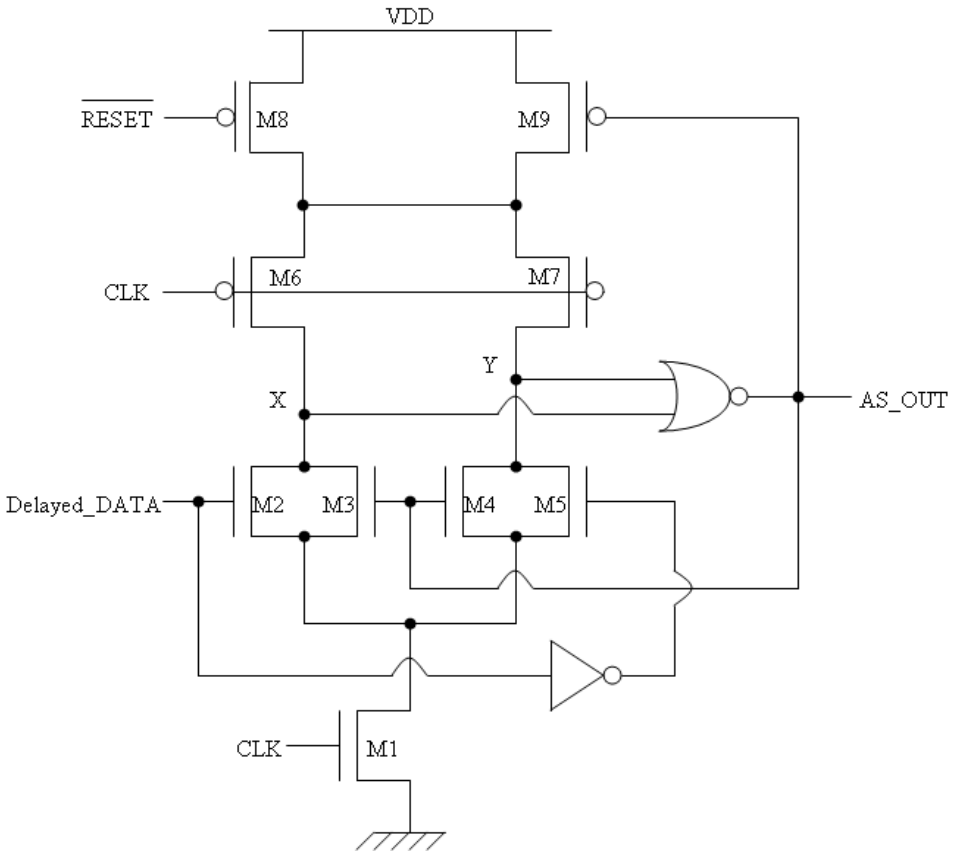


Figure 4.7. Stability checker architecture with on-retention logic.

During CLK low state, and considering that AS_out signal is low, X and Y nodes are pulled up (making AS_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 X or Y changes to low. If, during the high state of the CLK, a transition in Delayed_DATA occurs, the high X or Y node is pulled down by transistor M2 or M5, respectively, driving AS_out to go high. From now on, M9 transistor is OFF. Hence, X and Y nodes are not pulled up during CLK low state, unless the active low RESET signal is active. X and Y nodes remain low, helped by transistors' M3 and M4 activation during AS_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's active.

Moreover, only FF's internal clock signal is triggering the beginning of the observation interval, t_g. As mentioned, guard-band interval is the DE propagation delay, ultimately limited by half the clock cycle (when CLK signal is high).

Regarding the SC reliability, several factors like the input clock frequency, metastability and crosstalk, are all potential causes for sensor's malfunctioning.

Low frequency clock signals may affect the dynamic logic reliability. X or Y nodes may discharge during clock high state, due to drain current's leakage in transistors M2, M3, M4 and M5, and set the sensor's output high, resulting in a false-positive error. For high frequency clock signal, with discharge times extremely shorts, X or Y nodes voltage drop are negligible. However, below a certain clock frequency the discharge time is high enough to turn X or Y nodes low. It is therefore important to identify the minimum operation clock frequency to guarantee correct sensor operation. As an example, from simulations, 16Mhz was obtained as the minimum clock frequency for 65nm Berkeley PTM technology.

Metastability problems may occur if the sensor's output stuck at mid supply-voltage ($V_{dd}/2$), resulting in an undefined logic level for the connected digital circuitry. This may occur when delayed data signal level transitions occur with long rise/fall times. Simulation analysis, later presented in this work, reveals no metastability problem for the AEP-FF.

Regarding crosstalk effects, false detections could be triggered by the SC due to interferences during the positive clock half-cycle. To reduce crosstalk signal integrity problems in the DE and in the SC modules, especially when CLK is high and the sensor is enabled, these blocks, or the AEP-FF itself, should be designed with additional fault-tolerant design techniques, like the use of guard-rings.

4.3 CHARACTERIZATION

To characterize the AEP-FF, several simulations were performed using HSPICE™ and Agingcalc software applications. HSPICE™ is a commercial version of SPICE software, developed by Synopsys. SPICE is the acronym for *Simulated Program with Integrated Circuits Emphasis*, a software tool developed in 1972 at the University of California Berkeley, for analog and digital electronic circuits simulation and analysis, capable to perform DC, AC, transient, pole-zero, distortion, sensitivity or noise analysis. Since the mid-1960's, spice simulations are part of nearly every integrated circuit development [78]. Agingcalc is a software tool designed to analyze and predict digital circuit's aging induced by NBTI. This program does evaluate how individual transistors threshold voltages are affected with time, calculate circuit paths delays, find which FF are critical memory elements, and generates SPICE netlists for different aging moments in time. Agingcalc development started in 2010 at University of Algarve as part of Jackson Pachito's master studies, with the support of Prof. Jorge Semião, was released in 2011 and is currently under continuum development by the former [43].

On the following pages the results from several simulations will be exposed regarding the analysis of the new sensor's characteristics, with focus on FF setup and hold time measurements, sensor's guard-band measurements and behavior against PVTA variations, stability checker output characteristics and response time, and sensor's output stability.

On the subsequent section benchmark test circuits will be used to demonstrate the AEP-FF applicability and functionality, showing how its detection capability fits under different conditions, in particular under process variations.

All the simulations used 65nm Berkeley PTM technology models for p-channel and n-channel MOSFET transistors. The spice and verilog program files are in the attached CD together with the plots source values, saved in a spreadsheet.

This section uses two sets of parameters (set 1 and set 2) listed in Table 4.1 and Table 4.2, respectively, with different nominal conditions operation parameters and different DE circuit transistors channel length and width dimensions. The motive for this is linked with the time when simulations were performed. The first simulations were done before the Agingcalc program was ready, and used the first set of parameters. After Agingcalc development I choose to use a new set of parameter values to reflect more realistic working conditions. The first simulations were not repeated with the last set due to lack of time, and their use will not affect the analysis of the sensor's operation and behavior. However attention should be taken when comparing results from different simulations.

Technology model	PTM 65 ηm
Supply voltage	0.8 V to 1.2 V
Temperature	20 ºC to 150 ºC
V_{thP}	-0.365 V
T_{oxP}, T_{oxnN}	1.85 ηm , 1.95 ηm
Nominal-case operation conditions (NC)	V_{DD} = 1 V and T = 27 ºC
Worst-case operation conditions (WCC)	V_{DD} = 0.8 V and T = 150 ºC

Table 4.1. Simulation parameters set 1.

Technology model	PTM 65 ηm
Supply voltage	0.8 V to 1.2 V
Temperature	20 ºC to 150 ºC
V_{thP}	-0.365 V
T_{oxP}, T_{oxnN}	1.85 ηm , 1.95 ηm
Nominal-case operation conditions (NC)	V_{DD} = 1.1 V and T = 110 ºC
Worst-case operation conditions (WCC)	V_{DD} = 0.8 V and T = 150 ºC

Table 4.2. Simulation parameters set 2.

4.3.1 AEP-FF SETUP AND HOLD TIMES

Setup and hold times measurements were done for the AEP-FF and for a standard FF, with the same TG-MSFF architecture, to measure their effective value and to evaluate how the added DE and SC modules affects a standard FF cell.

To measure the setup time (see Figure 4.8), a step signal (S_{IN}) was applied at the flip-flop input D, and swept in time until the setup time was violated and the output (Q) stop updating its value to 1. At that moment the setup time ($\mathbf{t_{setup}}$) was measured as the time between the S_{IN} rising edge and the clock's rising edge.

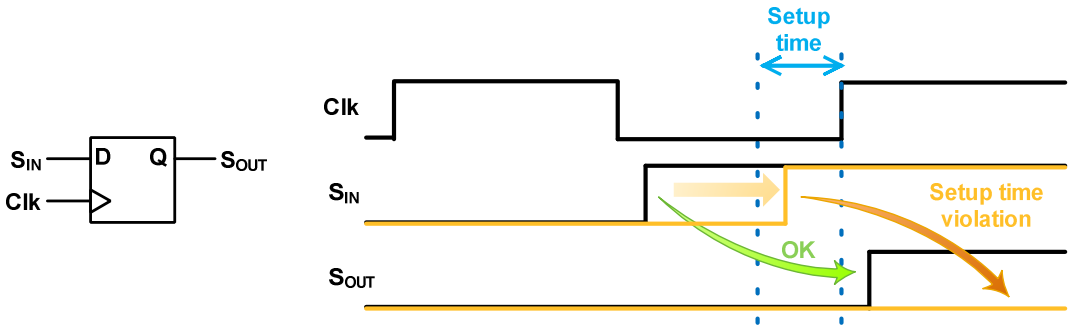


Figure 4.8. Flip-flop setup time measurement

To measure the hold time (see Figure 4.9) a pulse was applied at the flip-flop input D, starting exactly at t_{setup}. The pulse width was then reduced until the hold time was

violated and the output (Q) stop updating its value to 1. At that moment the hold time ($\mathbf{t_{hold}}$) was measured as the time between the clock's rising edge and the S_{IN} falling edge.

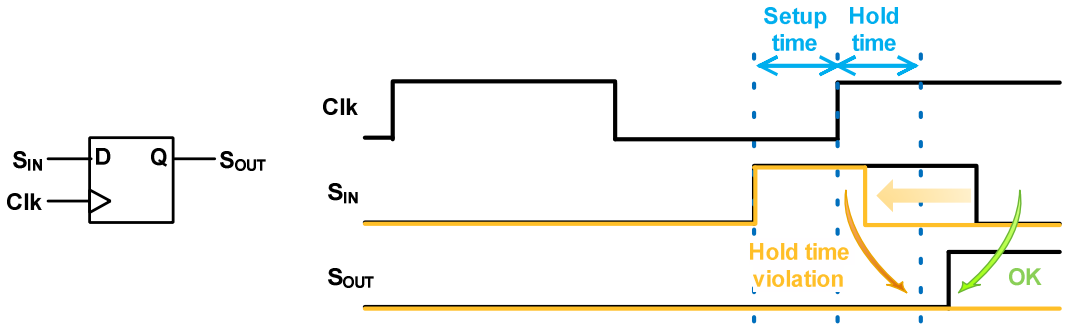


Figure 4.9. Flip-flop hold time measurement

The results were practically equal for both flip-flops with $t_{setup}+t_{hold}$= 52ps, showing no significant difference in these parameters when the aging sensor modules are connected to the FF circuit.

The simulations were performed in NC (Nominal Conditions), with V_{DD}=1V and T=27ºC.

4.3.2 AEP-FF GUARD-BAND

Ultimately, the AEP-FF guard-band is set by the time it takes for the data signal to go from the AEP-FF input D to the DE output, plus the time it takes for the SC to analyze the signal. In this chain, the DE is the main contributor for this delay time and the first analysis on the guard-band interval were done at this element. The first simulation test was made to analyze how each DE element circuit would behave against temperature, power supply voltage and NBTI induced PMOS transistors threshold voltage variations (VTA). Simulations were performed by applying a step signal at AEP-FF D input and measure the propagation time from DE_IN to DE_OUT (see Figure 4.10), with VTA variations, according to configuration set 1.

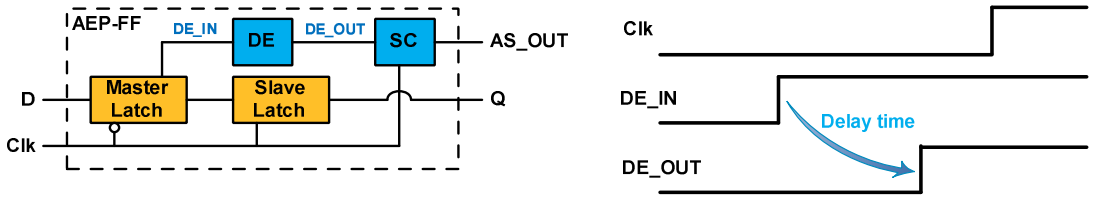


Figure 4.10. DE delay measurement configuration.

The results, plotted from Figure 4.11 to Figure 4.19, show that all DE circuits increase their delay time exponentially under VTA degradation (lower supply voltage, higher temperature and higher threshold voltage). DE_H is the more sensitive to temperature and aging degradation with up to 179% variation when temperature change from 20ºC to 150ºC, with V_{DD}=0.8V and no aging, and up to 31% delay variation for 15% increase in threshold voltage, with V_{DD}=0.8V and T=150ºC. Voltage variation causes similar delay behavior between the three DE, though it is visible higher delay variation at normal temperature (27ºC) for the DE_M and DE_H. From an application point of view, the absolute delay time differences between each delay element (19ps, 37ps and 125ps for DE_L, DE_M and DE_H respectively, under NC and nominal V_{thp}) demonstrate the great versatility of this sensor to be inserted in a broad range of circuits with different timing performances.

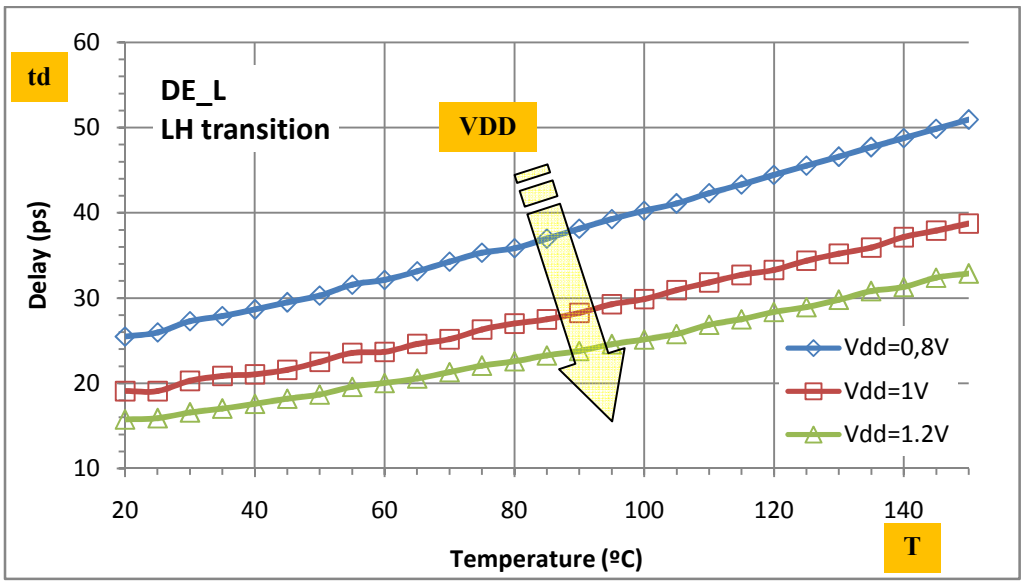


Figure 4.11. AEP-FF guard-band dependence on Temperature variations, with DE_L.

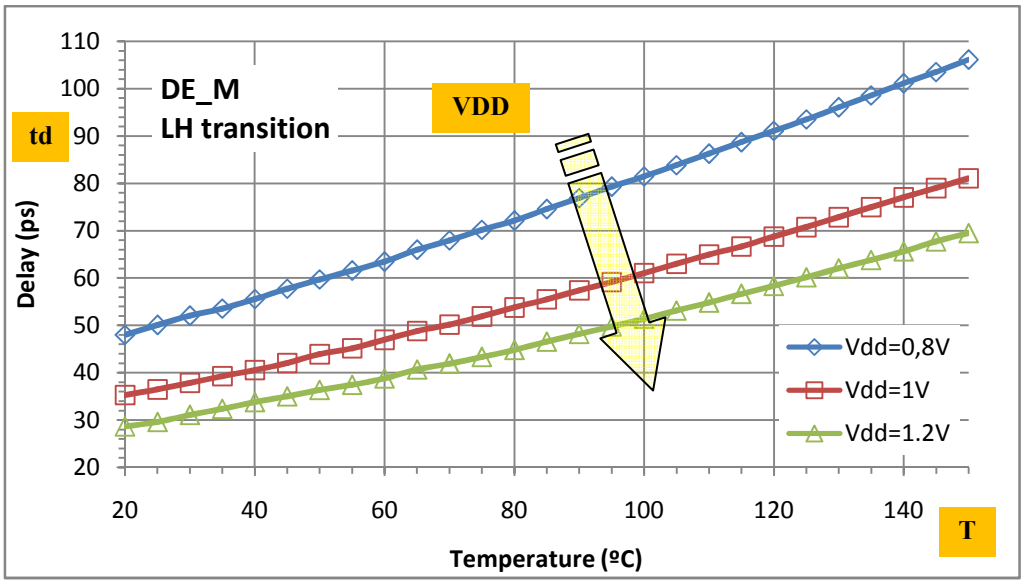


Figure 4.12. AEP-FF guard-band dependence on Temperature variations, with DE_M.

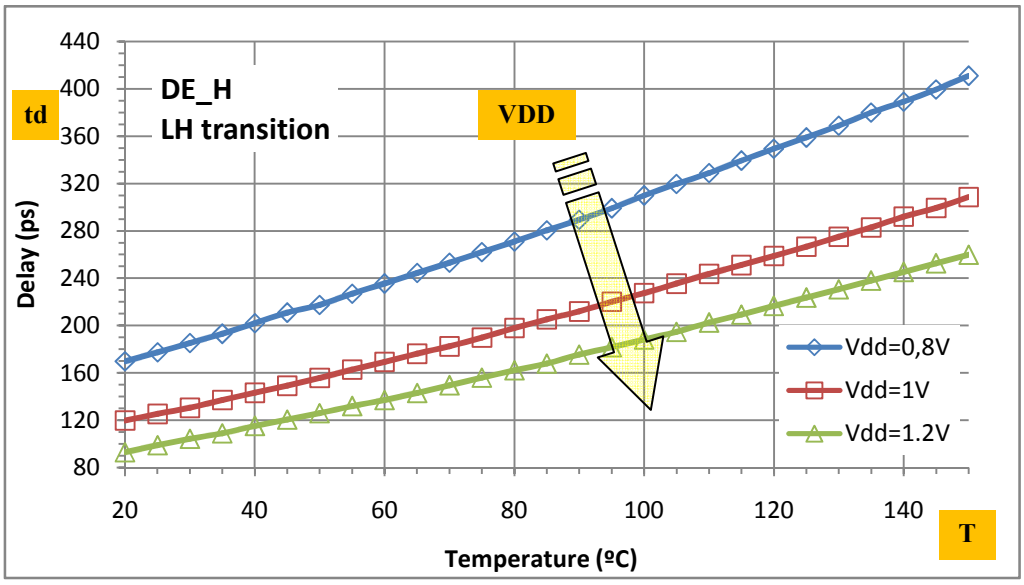


Figure 4.13. AEP-FF guard-band dependence on Temperature variations, with DE_H.

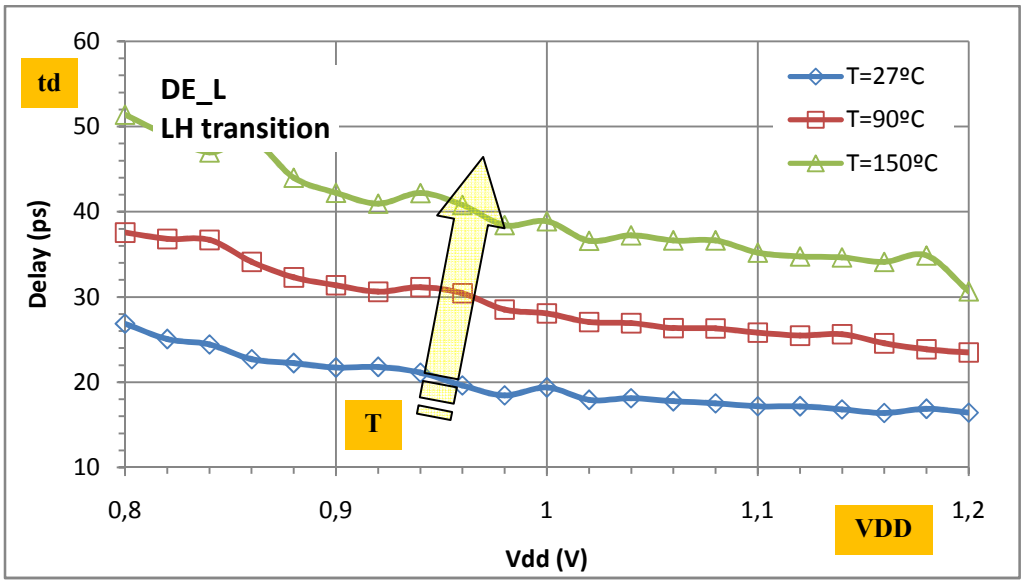


Figure 4.14. AEP-FF guard-band dependence on Power-Supply Voltage variations, with DE_L.

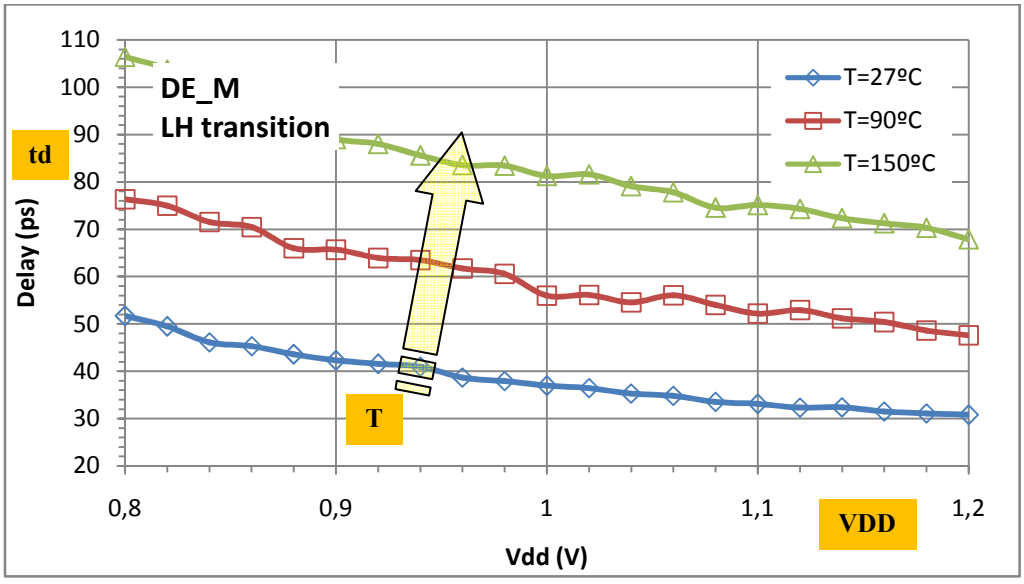


Figure 4.15. AEP-FF guard-band dependence on Power-Supply Voltage variations, with DE_M.

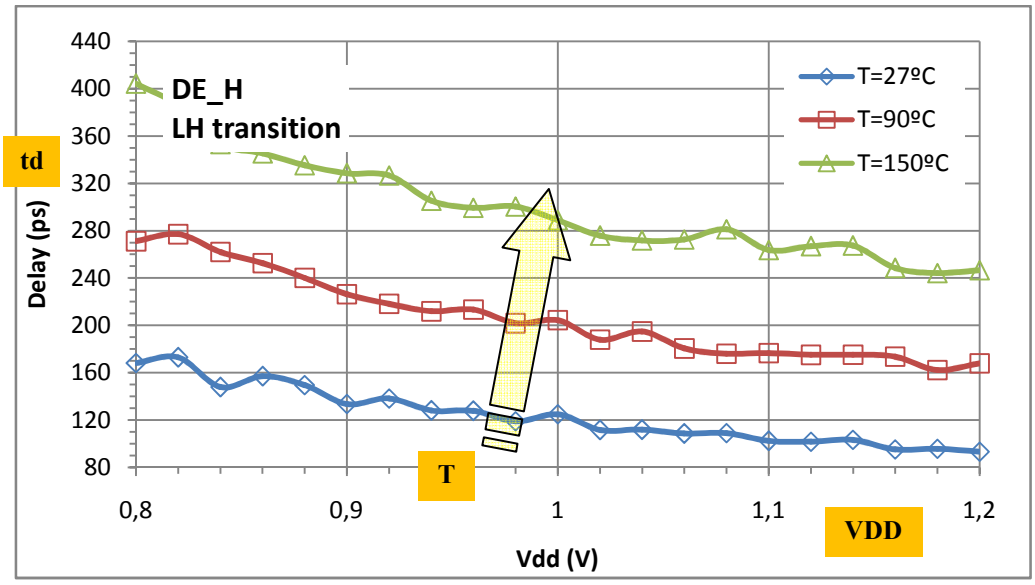


Figure 4.16. AEP-FF guard-band dependence on Power-Supply Voltage variations, with DE_H.

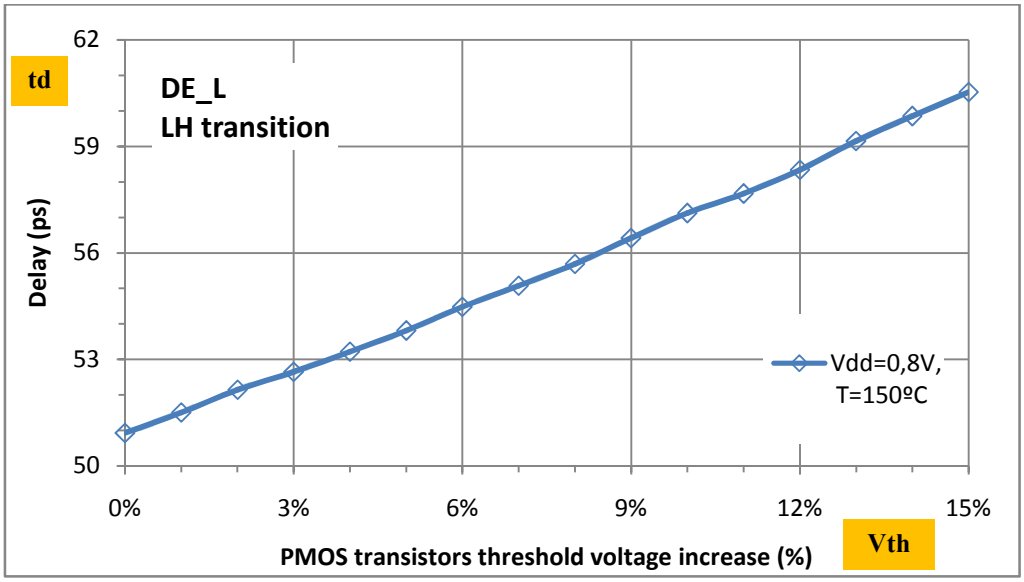


Figure 4.17. AEP-FF guard-band dependence on Vth variations, with DE_L.

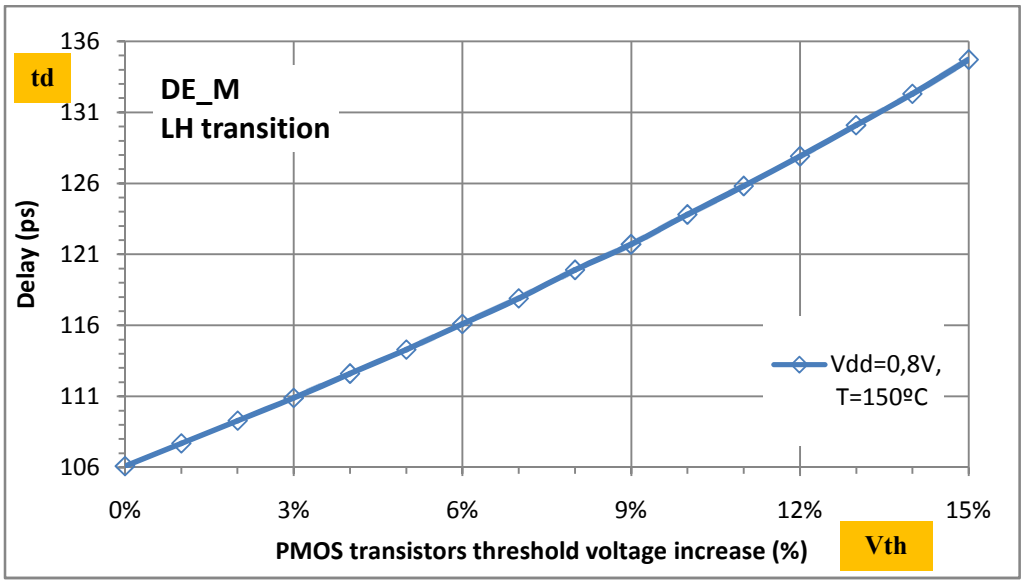


Figure 4.18. AEP-FF guard-band dependence on Vth variations, with DE_M.

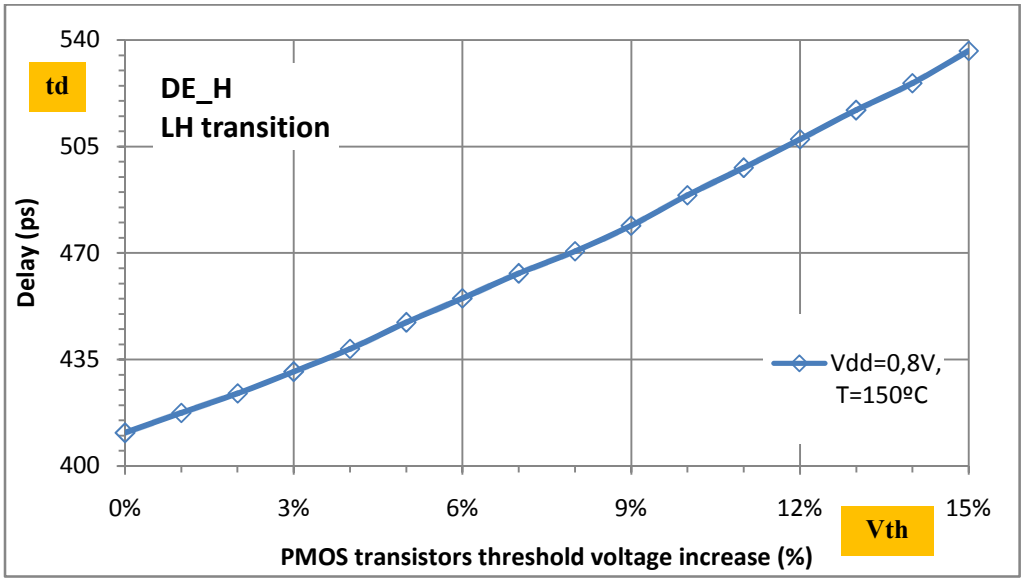


Figure 4.19. AEP-FF guard-band dependence on Vth variations, with DE_H.

Next, DE delay time dependence on process variations was analyzed performing delay measurements in 250 MC simulation runs, with 3σ equal to 30% variation from the nominal value, in transistor channel length, oxide thickness and threshold voltage, for test set 2, in WCC. The results, depicted in Figure 4.20, Figure 4.21 and Figure 4.22, reveal large delay time spans with process variations. Nominal values occur 13.6%, 17.6% and 14% of the time, for DE_L, DE_M and DE_H respectively.

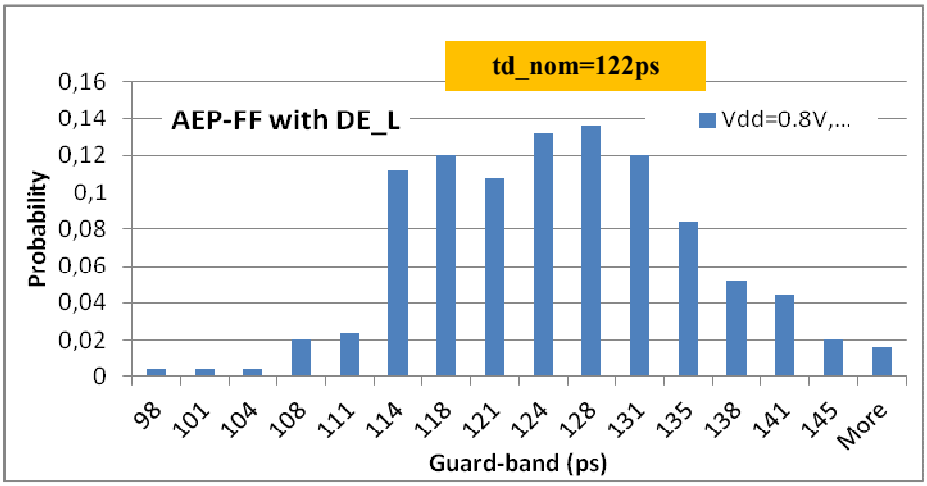


Figure 4.20. DE_L delay time dependence on process variation.

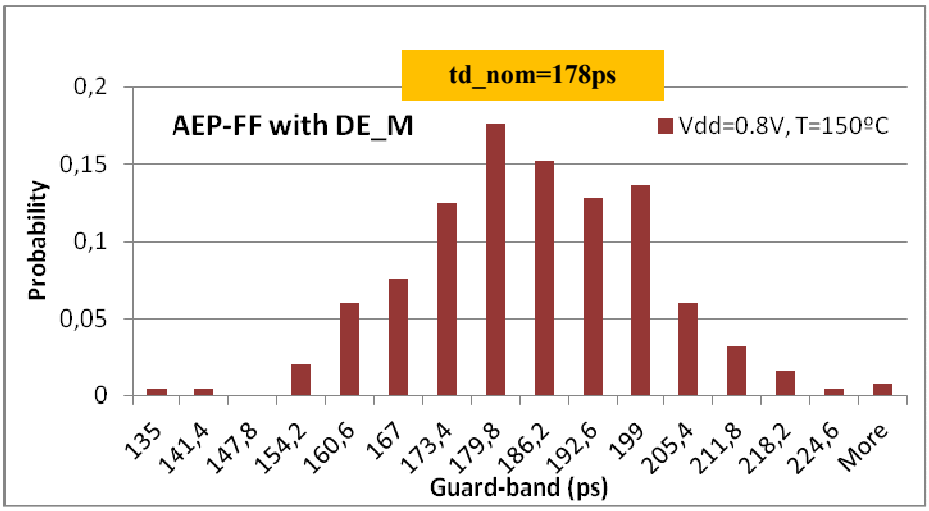


Figure 4.21. DE_M delay time dependence on process variation.

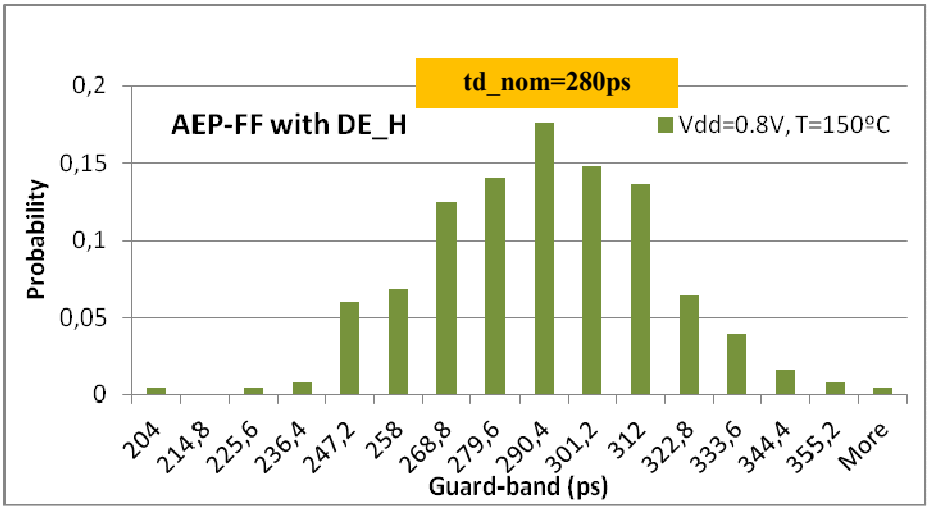


Figure 4.22. DE_H delay time dependence on process variation.

To analyze how the guard-band would age with operation time, a third set of simulations were performed on the AEP-FF. These were done using the parameters set number 2, to get the ideal guard-band interval for a working operation of 12 years considering 100Hz operating frequency. The measurements were performed applying a step signal at AEP-FF D input and measuring the propagation time from D to DE_OUT (see Figure 4.23), with aging variation under NC and WCC.

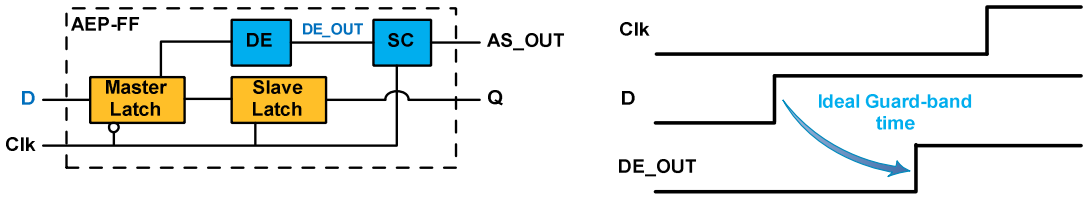


Figure 4.23.Ideal AEP-FF guard-band time measurement configuration.

Results show that guard-band increases logarithmically with time (Figure 4.24 and Figure 4.25), once more reveling the sensors sensitivity enhancement with aging. In nominal conditions (V_{DD}=1.1V and T=110ºC), guard-band time evolve from 68ps to 75ps (10%) with DE_L, 98ps to 115ps (17%) with DE_M, and 150ps to 176ps (17%) with DE_H, over 12 years operation time. In WCC, guard-band evolve from 122ps to 143ps (17%) with DE_L, 179ps to 229ps (28%) with DE_M and 280ps to 357ps (28%) with DE_H. These results show a strong guard-band time and aging dependence on the operation conditions. In WCC aging is accelerated, for example, with DE_H, the guard-band time increases 17% over 12 years operation in NC and 28% in WCC.

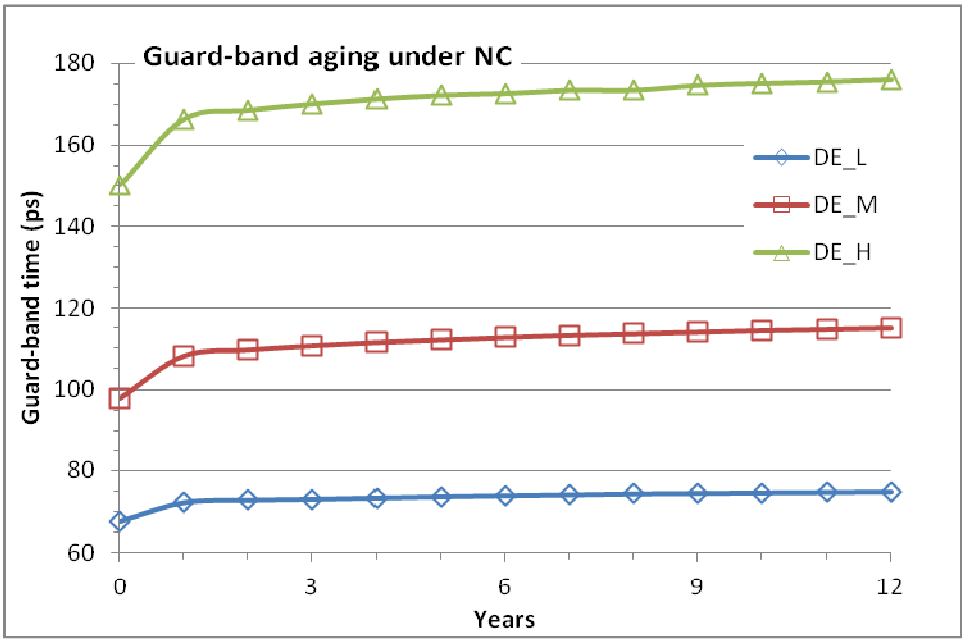


Figure 4.24.AEP-FF guard-band aging under NC.

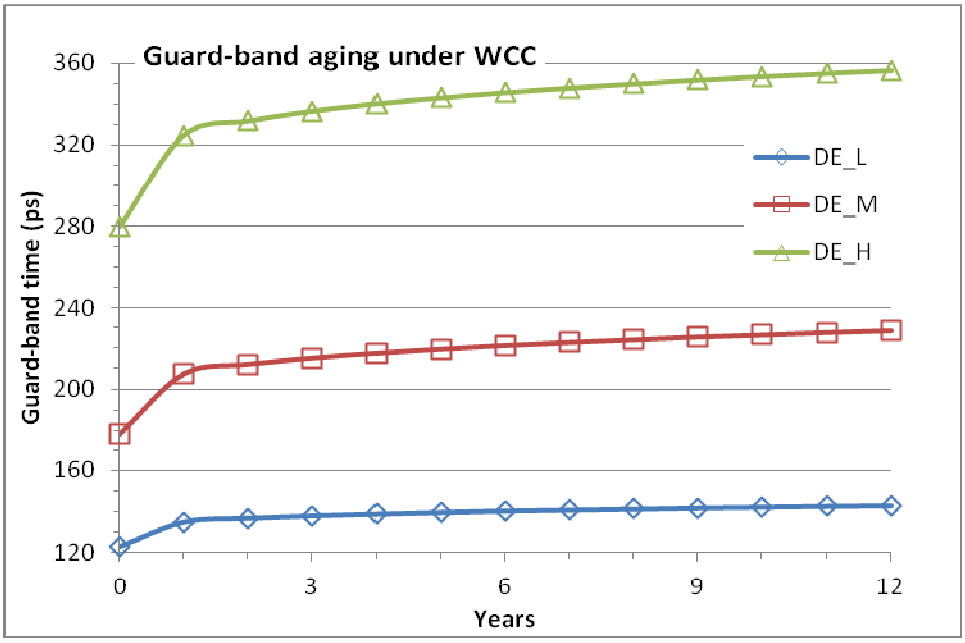


Figure 4.25.AEP-FF guard-band aging under WCC.

The ideal guard-band time measurement is a quick way to characterize a specific AEP-FF guard-band, being very useful at the design stage, when the AEP-FF is being tuned to fit the target circuit specifications. However, the ideal guard-band does not represent the real value when the sensor is in use, due to the time uncertainties in the SC response time. The method to get the effective sensor's guard-band time is a much slower process, done by applying a step signal to the AEP-FF D input and delaying

this signal until the sensor's output (AS_OUT) goes high (see Figure 4.26). The effective guard-band time is the time between the step signal transition, at AEP-FF D input, and the clock signal rising edge.

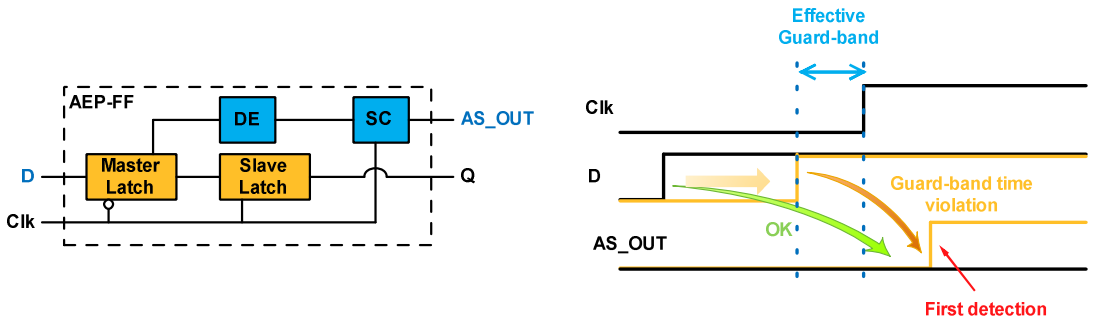


Figure 4.26.AEP-FF effective guard-band time measurement configuration.

The absolute measured values are plotted in Figure 4.27 and Figure 4.28. The effective guard-band times are smaller than the ideal ones, as expected, due to the time it takes for the SC to become active after clock signal rising edge. However, this is not considered a limitation, but instead a sensor's characteristic, because there is no signal defining the guard-band value. The guard-band interval is virtual and defined by construction and by sensor design. The value will differ for every local sensor, according to the PVTA variations applied locally.

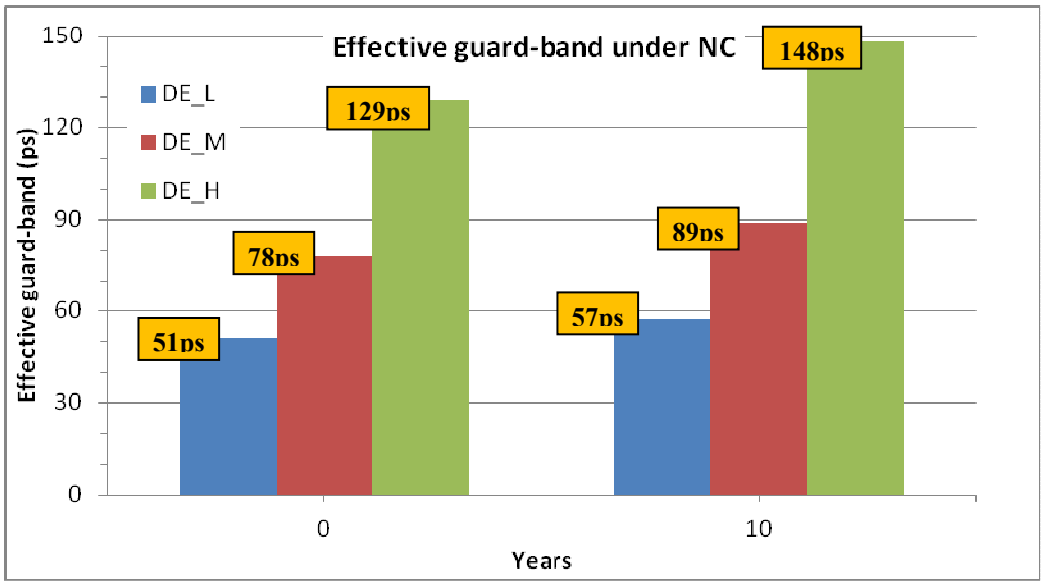


Figure 4.27. AEP-FF effective guard-band under NC

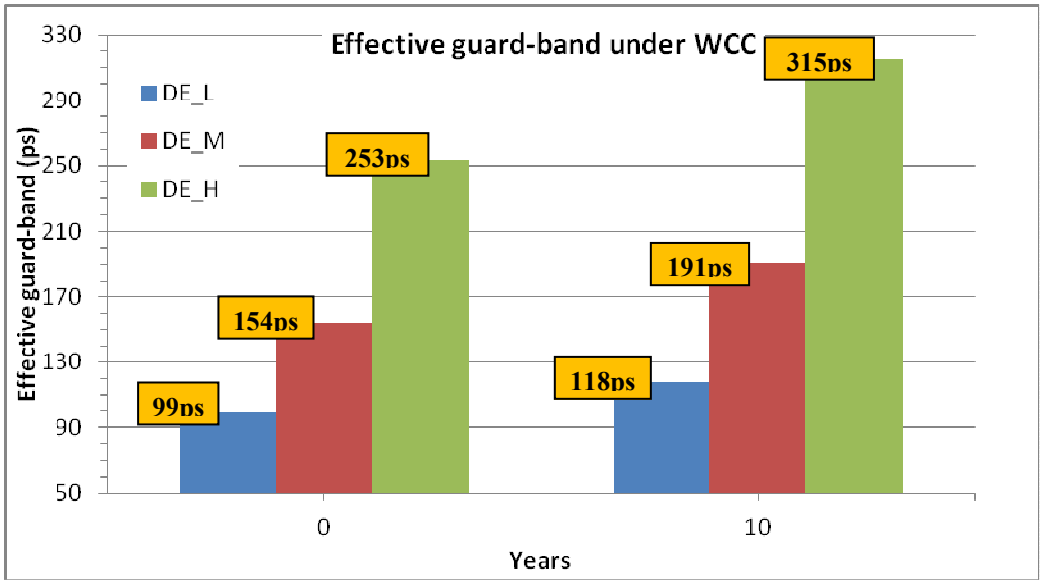


Figure 4.28. AEP-FF effective guard-band under WCC.

From the effective guard-band measurements it is possible to define target circuits for each DE element, based on the circuits operating frequency. For example, if the guard-band time for a critical path under monitoring represents 20% of total propagation time, the maximum clock period for the circuit would be given by (see Figure 2.19 and Figure 4.29):

$$f_{CLK_{max}} = \frac{0.2}{t_{gb}} \quad (3)$$

Where t_{gb} is the guard-band time. The minimum clock period must include F1 flip-flop's update time (t_{pcq}), critical path's propagation time (t_{pd}), the clock period skew (t_{skew}), the AEP-FF setup time (t_{setup}) and a minimum time slack to accommodate signal propagation variations (t_{slack}). In this case, the AEP-FF is adequate to monitor circuits typically up to 1.695 GHz, 1.05GHz and 635MHZ, when using DE_L, DE_M and DE_H, respectively (see Figure 4.30), operating under WCC for 10 years.

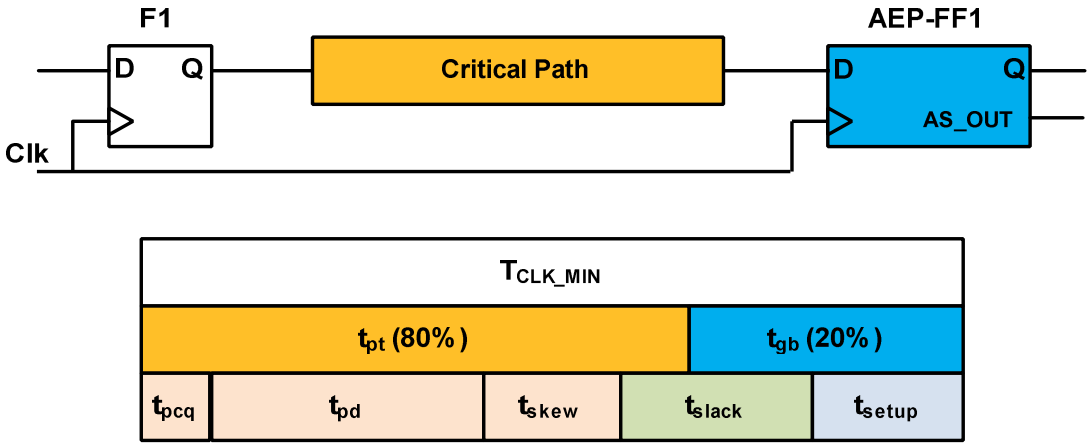


Figure 4.29.Target circuit's minimum clock period components.

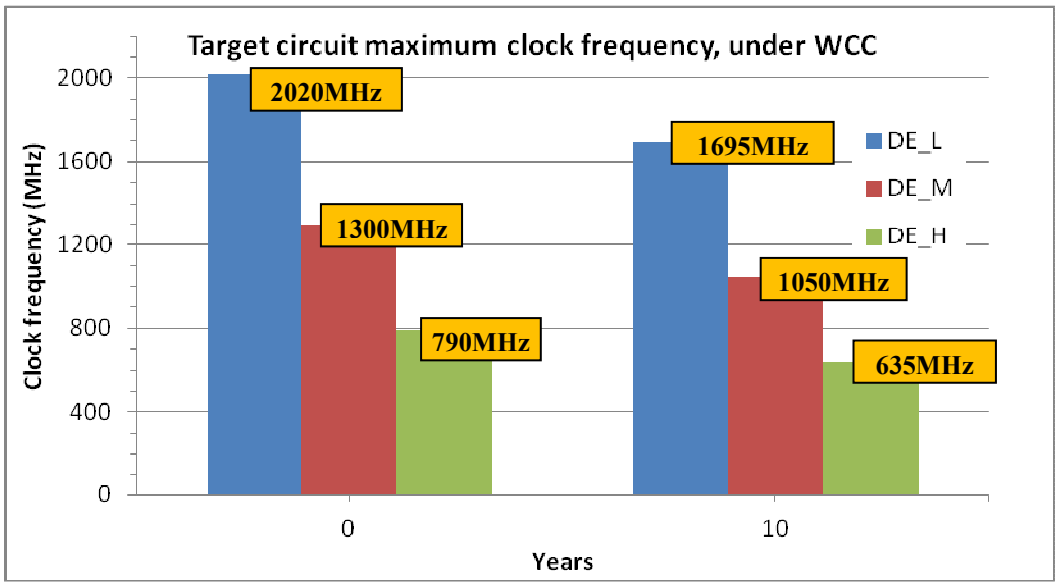


Figure 4.30.Target circuit maximum frequency.

For proper error prediction it is imperative that the sensors guard-band have a longer time than the FF setup time. No errors should occur in the FF captured data in the exact moment that a delayed signal enters the sensors guard-band time window, triggering a guard-band violation. However, if the time degradation keeps growing, the delayed signal will eventually be high enough to violate the FF setup time, resulting in a delay fault. To understand the AEP-FF behavior vs data signal delay, the FF output was also monitored during the previous simulations. After guard-band time measurement, the signal delay was increased until the FF output stop updating its values (see Figure 4.31).

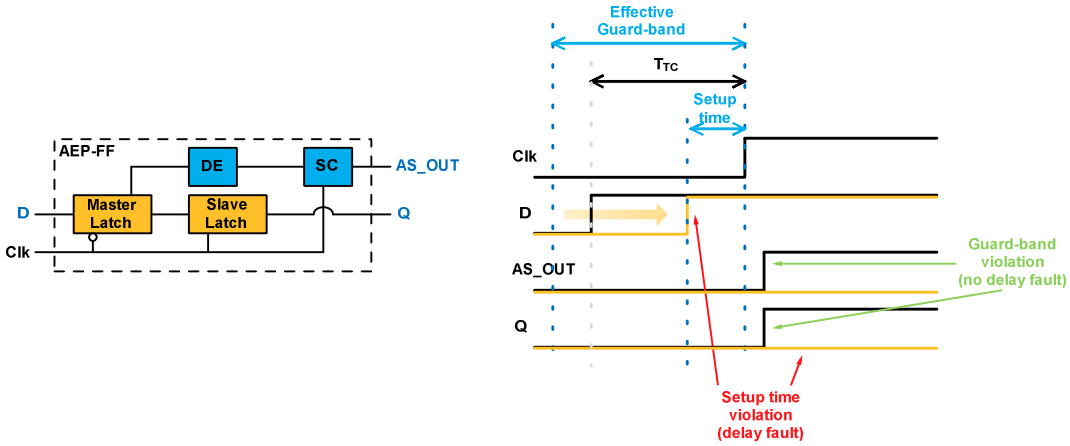


Figure 4.31. AEP-FF delay fault analysis configuration.

Results show that error-prediction continues until the fault occurs (Figure 4.32 and Figure 4.33)), which indicates that there is no sensor's detectability loss at the end of the guard-band interval, nor false detections at sensor's output (this is a major improvement over previous sensors). At year 0, in WCC, delay faults start to occur for data transitions between 9ps and 0ps before the clock rising edge (T_{TC}). It is important to understand that the aging sensor alone does not avoid delay faults, and posterior actions on the circuit operation need to be done as soon as possible after the first guard-band violation detections. A possible solution is to reduce the target circuit clock frequency or increase the circuit's power supply voltage.

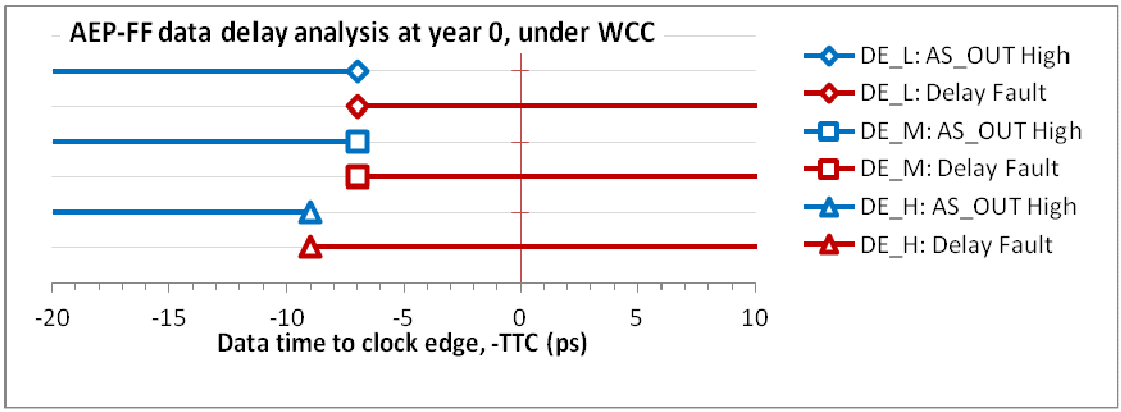


Figure 4.32. AEP-FF data delay analysis at year 0, under WCC (clock rising edge occurs at 0ps).

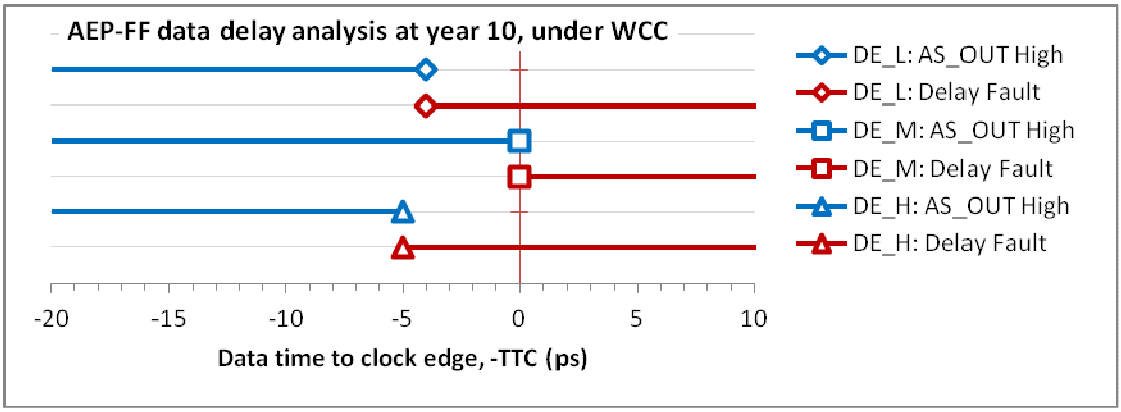


Figure 4.33. AEP-FF data delay analysis at year 10, under WCC (clock rising edge occurs at 0ps).

4.3.3 SC ANALYSIS

Tests made to the SC focused on: understand how input signal rise/fall time would affect the detection capability; find how much time it takes for a guard-band violation to be detected, when the clock signal goes high; identify any region were detection is not guaranteed; analyze the SC metastability, looking for $V_{DD}/2$ output locking; and analyze the SC output stability against false positives.

All these tests share a common simulation configuration, using configuration set 1 in WCC. A ramp signal was applied to the SC input and swept in time around the clock signal rising edge, repeating the sweeps for different rise/fall times (see Figure 4.34). Meanwhile, the SC output was monitored.

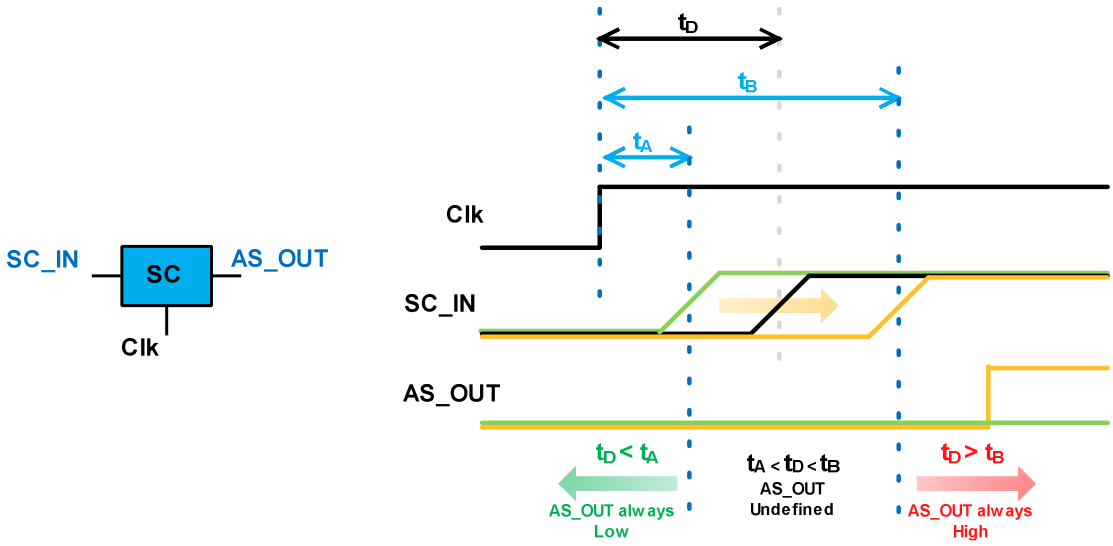


Figure 4.34.SC analysis setup.

From the simulations we can identify 3 detection zones (see Figure 4.34): (i) if the signal arrives (t_D) before t_A, AS_OUT will always be low and any signal variation at SC_IN is ignored; (ii) if the signal arrives after t_B, AS_OUT will always be high; (iii) if t_D is between t_A and t_B, AS_OUT logic level is undefined. The simulation results show that t_A and t_B are highly dependent on the input signal rise/fall time (t_R and t_F), with values changing more than 10ps for rise/fall times between 0.8ps to 400ps (see Figure 4.35), with HL transitions being detected later than LH transitions. It is important to highlight that detections occur later when the ramp time increases. As a result, the additional time t_A will add to the previously measured effective guard-band, increasing the sensor's sensitivity even further for slower signals (signal slope time increases when circuits get slower, e.g. due to PVTA variation effects)), which is again a very good result, when compared to previously published aging sensors, highlighting the robustness of the present solution.

Regarding the undefined region, it shows no significant changes and relationship with the input signal rise/fall time. The zone where the output could be undefined is very small, ranging from 0.1ps to 0.7ps (see Figure 4.36).

Any metastability problems should occur within t_A and t_B. With 0.1ps sweep time resolution no metastability problems were noticed in the simulations performed to the circuit and the SC never lock to an undefined logic value (near Vdd/2), being always at logic level high or low.

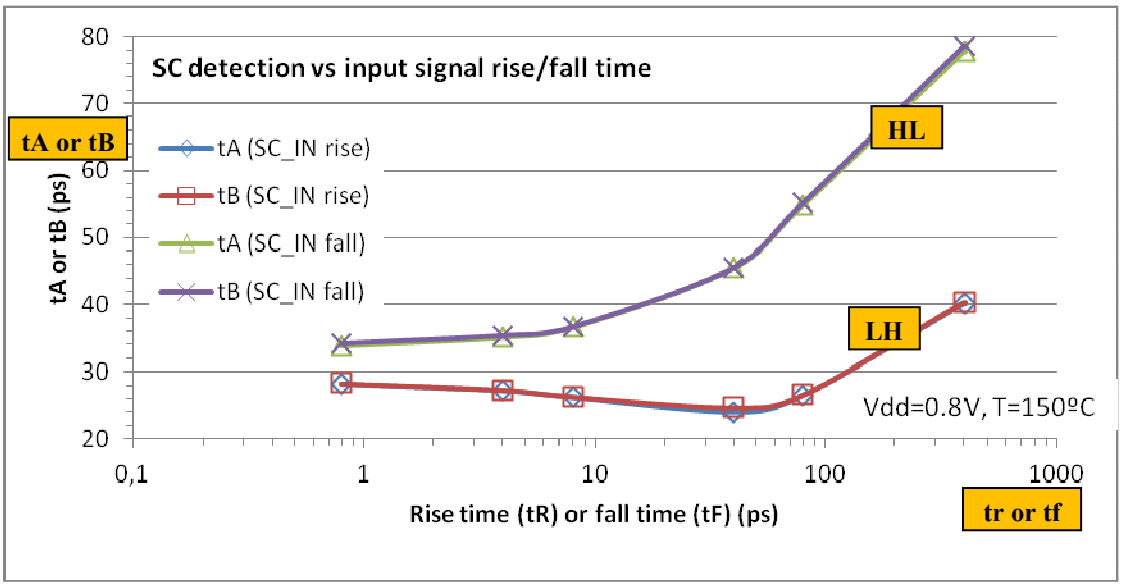


Figure 4.35. SC detection vs input signal rise/fall time, in WCC.

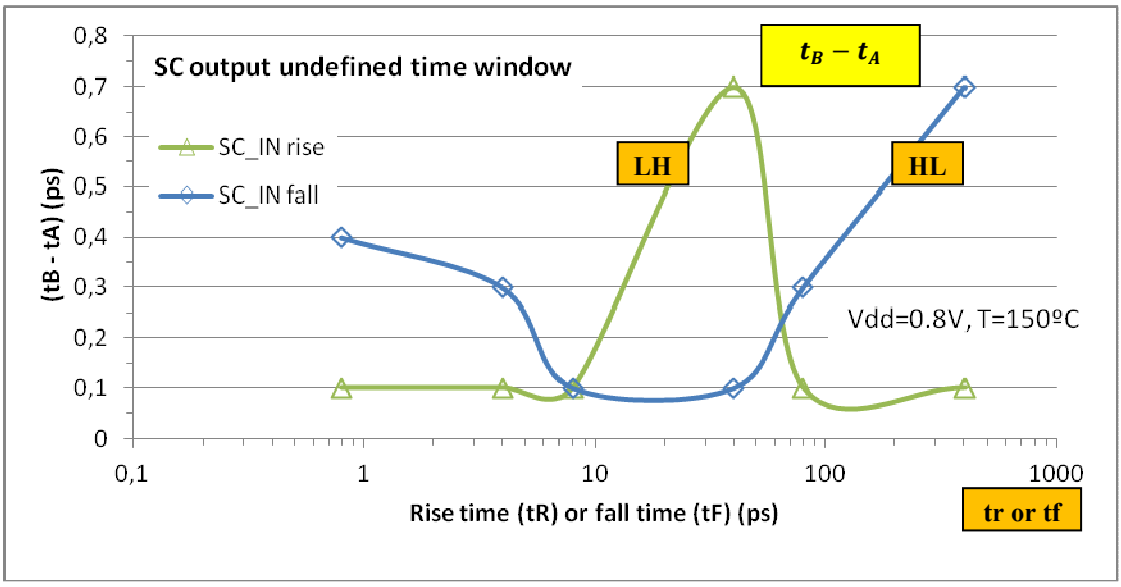


Figure 4.36. SC output undefined time window.

Regarding the SC output stability, simulations were performed with high clock periods up to 60ns with no false positives. For a reliable operation the SC dynamic logic, refresh rate should be at least 16.7MHz (f_{clk}> 16.7MHz), which is completely achievable for nanometer technologies with much higher clock rates.

4.3.4 DATA SIGNAL GLITCHES

Glitches on data signals that occur inside the guard-band interval may induce false detections on the aging sensor. The motivation to study how glitches affect the sensor is to analyze its robustness to cross-talk. Cross-talk analysis is very complex, require models that account for all parasitic capacitances and inductances between circuit traces and active elements, being very much dependent on the circuit layout. This turns general cross-talk simulation and analysis almost impossible, being only feasible for a specific circuit layout. However, for digital signals, cross-talk effects usually result in pulses or glitches on the victim lines. Considering the aging sensor, any connection may be affected by cross-talk. A solution for cross-talk is to design a guard-ring around the sensor circuit. This will protect the internal connections but will not avoid glitches coming from external lines, like the signal data line. In this way, a simple form to classify the AEP-FF with respect to cross-talk is to measure how long a glitch (pulse) should be to cause a false detection.

Simulations performed with the AEP-FF using the three DE circuits were done to find the minimum width for a glitch to be detected. The minimum values were measured with circuit operating in the best conditions to get the fastest response time to variations. The measurements were made with the circuit operating with VDD = 1.1V and T = 27ºC, with no aging, modeling glitches as positive pulses (low-high-low) or negative pulses (high-low-high) (see Figure 4.37).

Measurement results show that pulses whose widths are below 20.8ps are not detected by the sensor, for a positive pulse applied to AEP-FF with DE_L. With DE_M and DE_H, the minimum widths were 27.2ps and 38.3ps respectively. In this way, as long as a glitch width is shorter than the aforementioned values, that glitch will be ignored. Considering the high measured values, it is somehow safe to say that the sensor is quite insensitive to glitches.

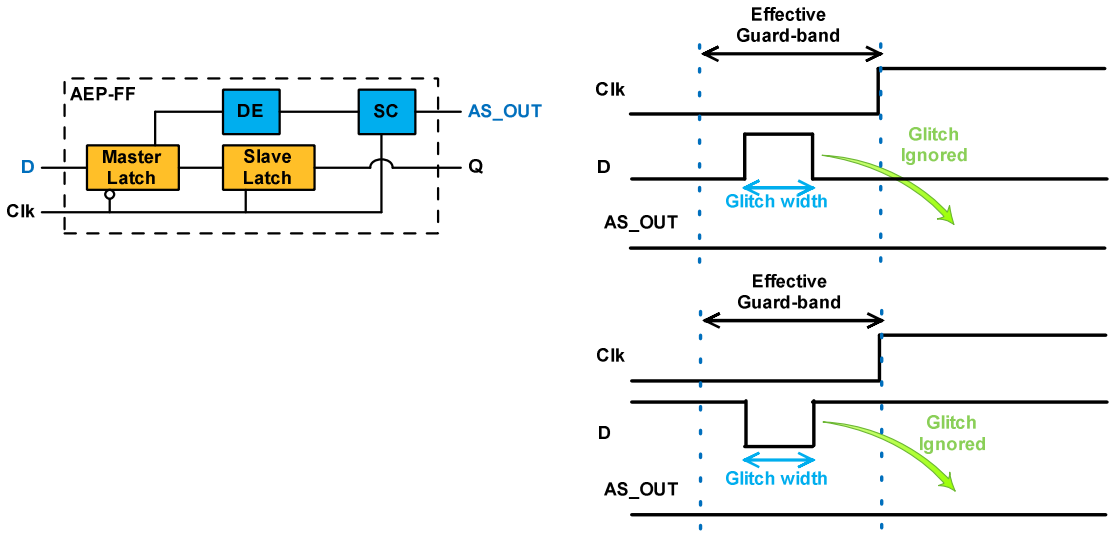


Figure 4.37. Glitch analysis configuration.

5. SIMULATION RESULTS

Simulation results are presented in this chapter for two benchmark circuits, to demonstrate the AEP-FF applicability. The first circuit is an inverter chain, chosen due to its simplicity, to demonstrate the AEP-FF detection capability and process variability dependence. Then a pipeline multiplier was used to demonstrate how several sensors in the same circuit may be used to monitor a complex circuit and detect delay degradation at different moments in time. At the end of the chapter, a comparison is made between AEP-FF characteristics and the previously developed sensors.

The simulations presented in this subsection used the parameters set in Table 5.1.

Technology model	PTM 65 ηm
Supply voltage	0.8 V to 1.2 V
Temperature	20 ºC to 150 ºC
V_{thP}	-0.365 V
T_{oxP}, T_{oxnN}	1.85 ηm , 1.95 ηm
Nominal-case operation conditions (NC)	V_{DD} = 1.1 V and T =110 ºC
Worst-case operation conditions (WCC)	V_{DD} = 0.8 V and T = 150 ºC

Table 5.1.Simulation parameters for benchmark circuits.

5.1 INVERTER CHAIN CIRCUIT:

The inverter chain circuit schematic is depicted in Figure 5.1. It is composed of one input and one output FF and 20 inverters connected in cascade. The 20 inverters chain will represent a critical path and the output FF is an AEP-FF.

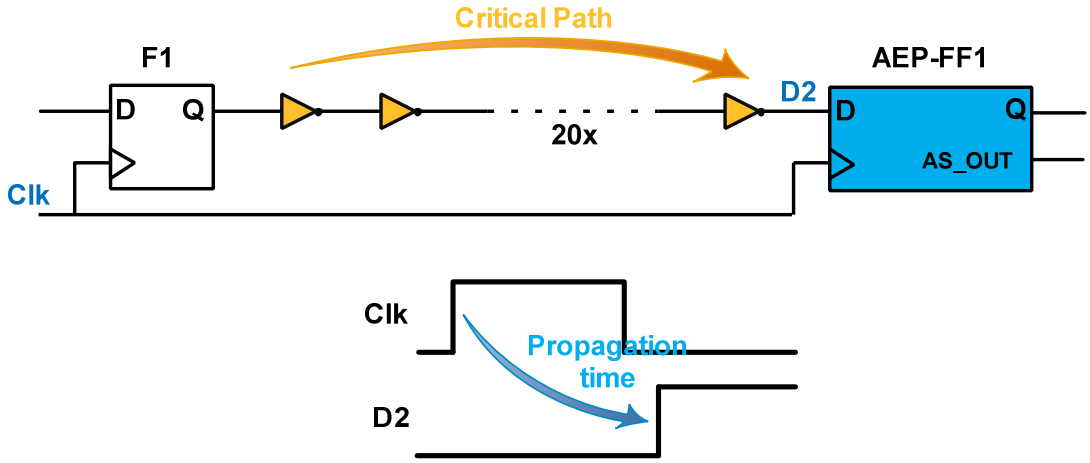


Figure 5.1.20 Inverter chain + AEP-FF schematic.

Initial propagation time measurements made to the inverter chain circuit with a regular output FF, instead of AEP-FF, revealed significant time degradation when the circuit is subjected to VTA variations (see Figure 5.2 and Figure 5.3). At year 0, in NC, nominal propagation time is 302.6ps, a value that almost doubles after 12 years of circuit operation under WCC, to 667.5ps.

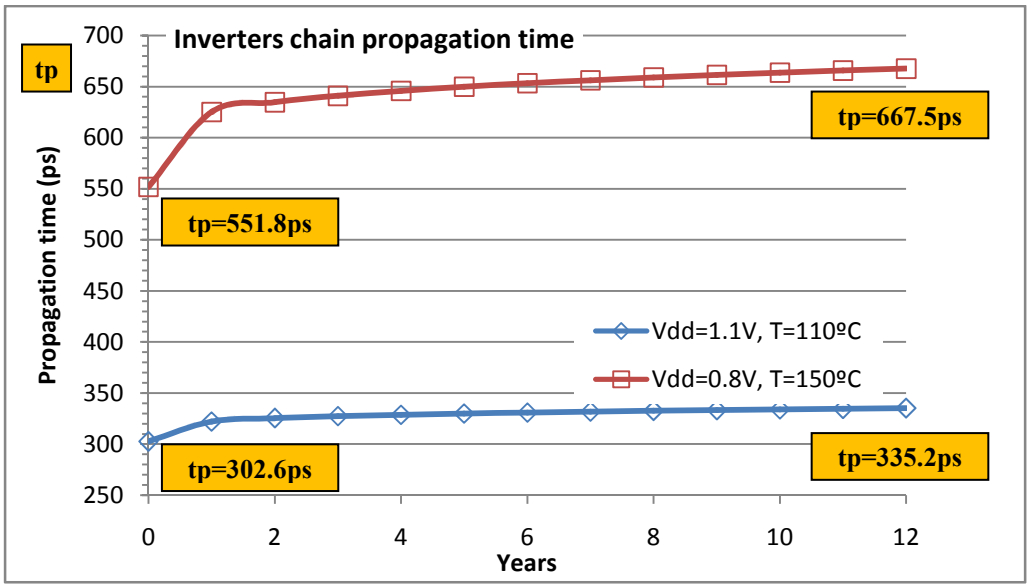


Figure 5.2. Inverter chain propagation time (with regular FF).

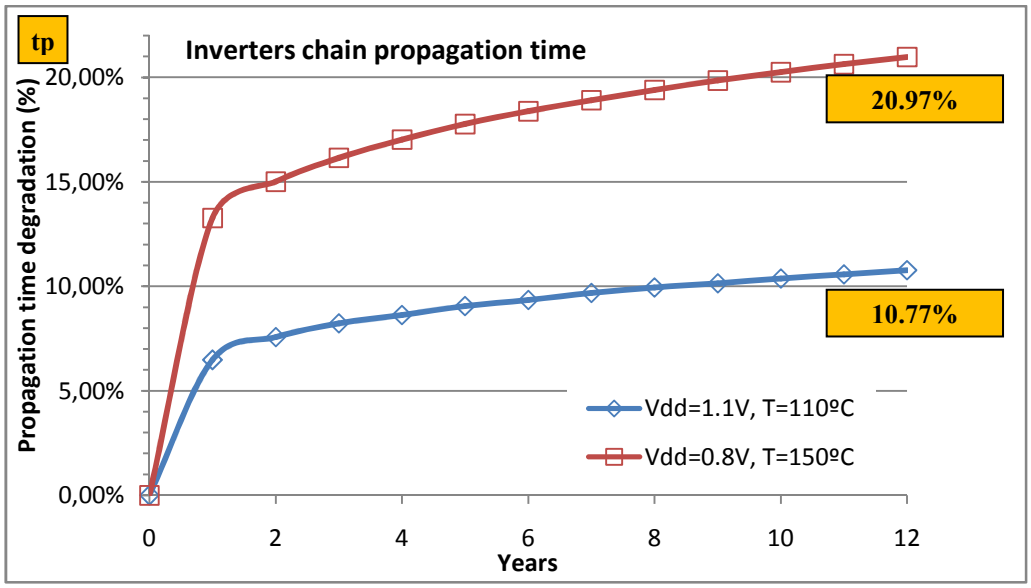


Figure 5.3. Inverter chain propagation time degradation (with regular FF).

Considering that the circuit's clock has 20% of its period reserved for the guard-band, then the DE effective guard-band should be tuned to values between approximately 75ps and 167ps (equations (4) and (5)). For these conditions, DE_M circuit was chosen as the best option to be used in the AEP-FF, with an effective guard-band ranging between 78ps and 191ps.

$$T_{clk-min-NC} = \frac{300ps}{0.8} = 375ps \quad \Longrightarrow \quad t_{gb-NC} = 0.2 * 375 = 75ps \tag{4}$$

$$T_{clk-min-WCC} = \frac{668ps}{0.8} = 835ps \quad \Longrightarrow \quad t_{gb-WCC} = 0.2 * 835 = 167ps \tag{5}$$

5.1.1 PERFORMANCE IMPACT

After the FF replacement by the AEP-FF, the previous measurements were repeated to evaluate if and how the sensor insertion would affect the signal path's performance, i.e. its propagation time. Measurements returned negligible performance impact, with propagation time differences inferior to 0.16% in NC.

These are expected results considering the AEP-FF architecture, where the aging sensor is connected to the FF internal nodes, isolated from the path nodes.

5.1.2 EFFECTIVE GUARD-BAND

Measurements made to the sensor's effective guard-band after insertion (see Figure 5.4), revealed values up to 19% higher than the ones obtained from individual AEP-FF analysis. This difference is expected because signals at the AEP-FF D input have higher rise/fall times than the test ones, resulting in added delay in the SC module. The effective guard-band measurements were done for data signals with fast, 1ps, rise/fall times, while measurements made at the AEP-FF D input returned values for signal rise time ranging from 128ps to 168ps.

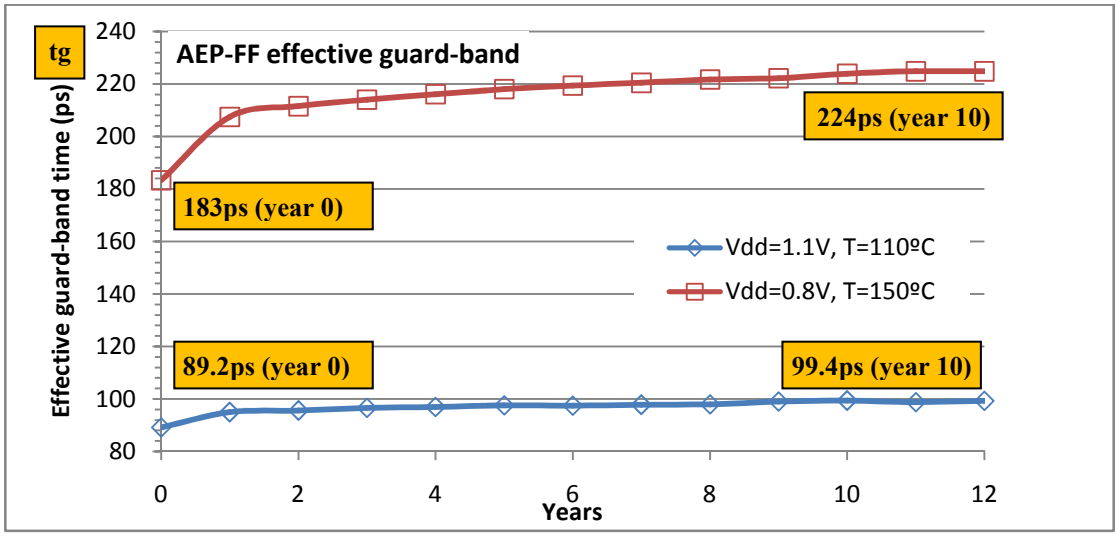


Figure 5.4. Inverter chain's AEP-FF effective guard-band .

5.1.3 MAXIMUM CLOCK FREQUENCY

Next, different clock frequencies were applied to the circuit for different VTA conditions, to test the aging sensor detection capability and find the maximum operation clock frequency that will avoid guard-band violation (Figure 5.5). VTA

variations were defined considering a 12 years circuit life span, operating under NC and WCC.

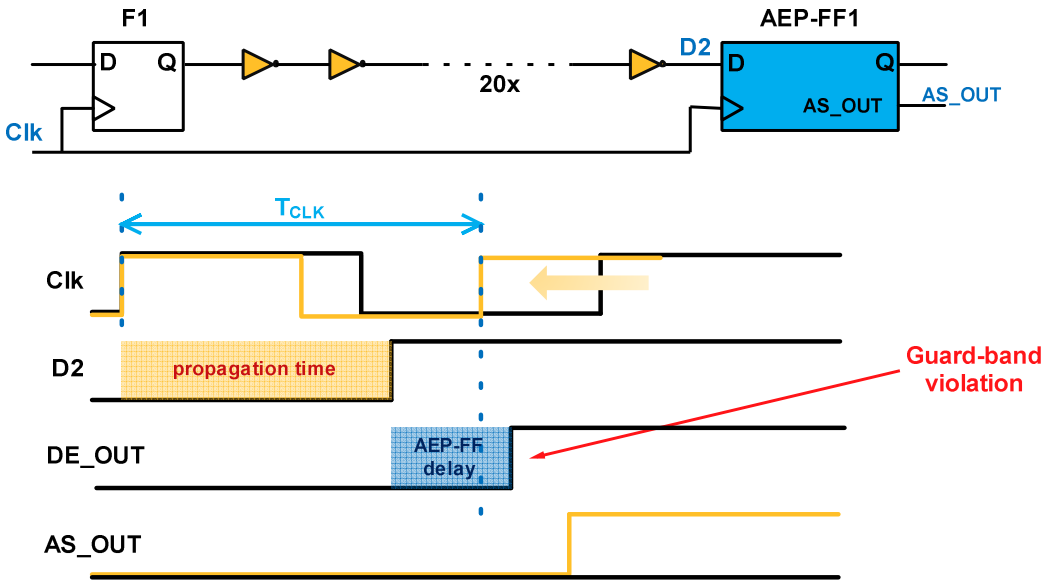


Figure 5.5. Inverter chain test configuration.

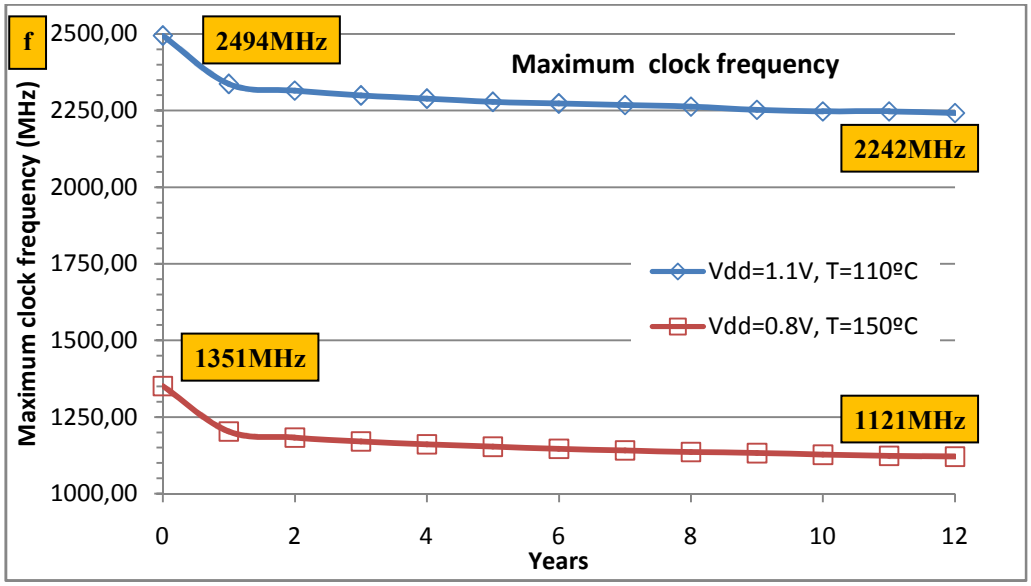


Figure 5.6. Inverter chain maximum clock frequency.

For this configuration up to 2.49GHz clock frequencies may be used in NC. However, this clock frequency is severely reduced due to aging and VT variations. The results, plotted in Figure 5.6, show clock frequencies going down up to 55%, from year 0 in NC (2.49GHz) to year 12 in WCC (1.12GHz). On a traditional circuit

design, the clock frequency would be limited to 1.12GHz to guarantee reliable circuit operation over the 12 years life span, without guard-band violation. Using the AEP-FF aided by self-healing mechanisms, like dynamic frequency scaling, it is possible to track system degradation over time and adjust its clock frequency to get the maximum performance every time, with assured reliability. For this example, performance gain could be up to 100%, with the target circuit adjusting its operating frequency dynamically between 2.29GHz and 1.12GHz, according to the operation and aging conditions.

5.1.4 MAXIMUM CLOCK FREQUENCY FOR AEP-FF OPERATION

The safety margin introduced by the guard-band timing is ultimately limited by delay faults and by the clock high time period. If no self-healing actions are taken after guard-band violation signaling, the path delay will keep growing and will eventually cause a delay fault. On the sensor side, the guard-band violation detections occur only if and when the delayed signal transitions arrive within the clock-high time window, at SC input. If transitions appear at the SC input latter, when the clock signal is low, the sensor will not detect it. Using the same analysis approach used before, to measure the maximum clock frequencies, it is possible to determine for what frequencies the sensor stop working. To perform this measurement, the clock frequency was further increased until the aging sensor output return to a low level state (Figure 5.7).

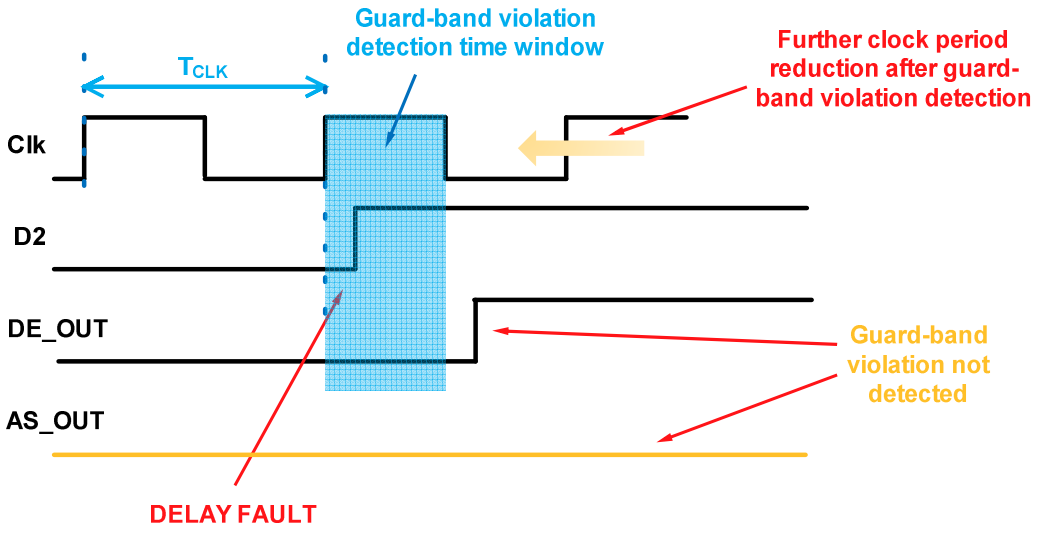


Figure 5.7.AEP-FF guard-band detection time window.

For this circuit, operating frequencies up to 3GHz are possible before the sensor stop signaling guard-band violations (Figure 5.8).

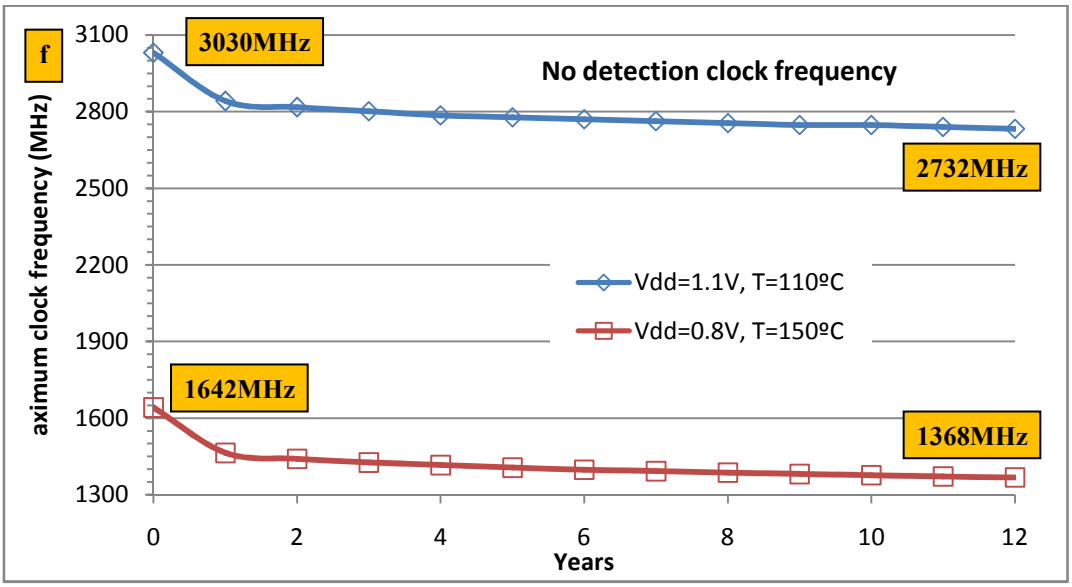


Figure 5.8.Inverter chain maximum clock frequency for AEP-FF operation.

5.1.5 CIRCUIT FAILURE PREDICTION

Finally, to demonstrate the sensor's functionality, the aging sensor and the inverters chain circuit were tuned for 10 years operation time, and the sensor output

was monitored along 20 years, with circuit operating in WCC, with a clock frequency around 1125MHz. Detections start to occur at the 11th year (Figure 5.9), as expected.

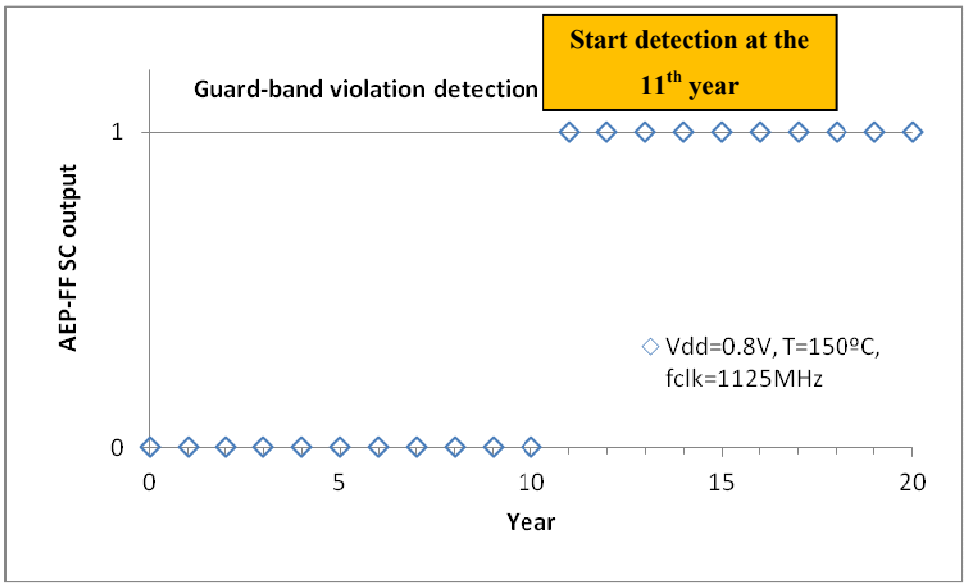


Figure 5.9. Inverter chain guard-band violation detection example.

5.1.6 FAILURE PREDICTION UNDER PROCESS VARIATIONS

To validate the circuit's operation under process variability 30 monte-carlo (MC) simulations were performed. Transistors length, oxide thickness and threshold voltage parameters were defined as random variables, with a Gaussian distribution, where each parameter's mean equal the nominal value and varies +-10% within 99.7% of the times (within 3σ[8]). Each MC run defines one circuit to be tested. For each MC run, the aging sensor output was monitored over a 20 years operation life span. The results show that process variability affects the circuit's performance significantly (Figure 5.10). Looking at the statistical analysis, plotted in Figure 5.11, only year 0 returned a no detection condition, and, within the 20 years simulated life span, 100% detection probability is never guaranteed in any year. This circuit behavior uncertainty due to manufacturing variability may be overcome with circuit binning,

[8] Three times the standard deviation.

classifying the different manufactured chips by performance/aging grades. Besides this, the AEP-FF sensor itself may be used for chip binning after manufacture.

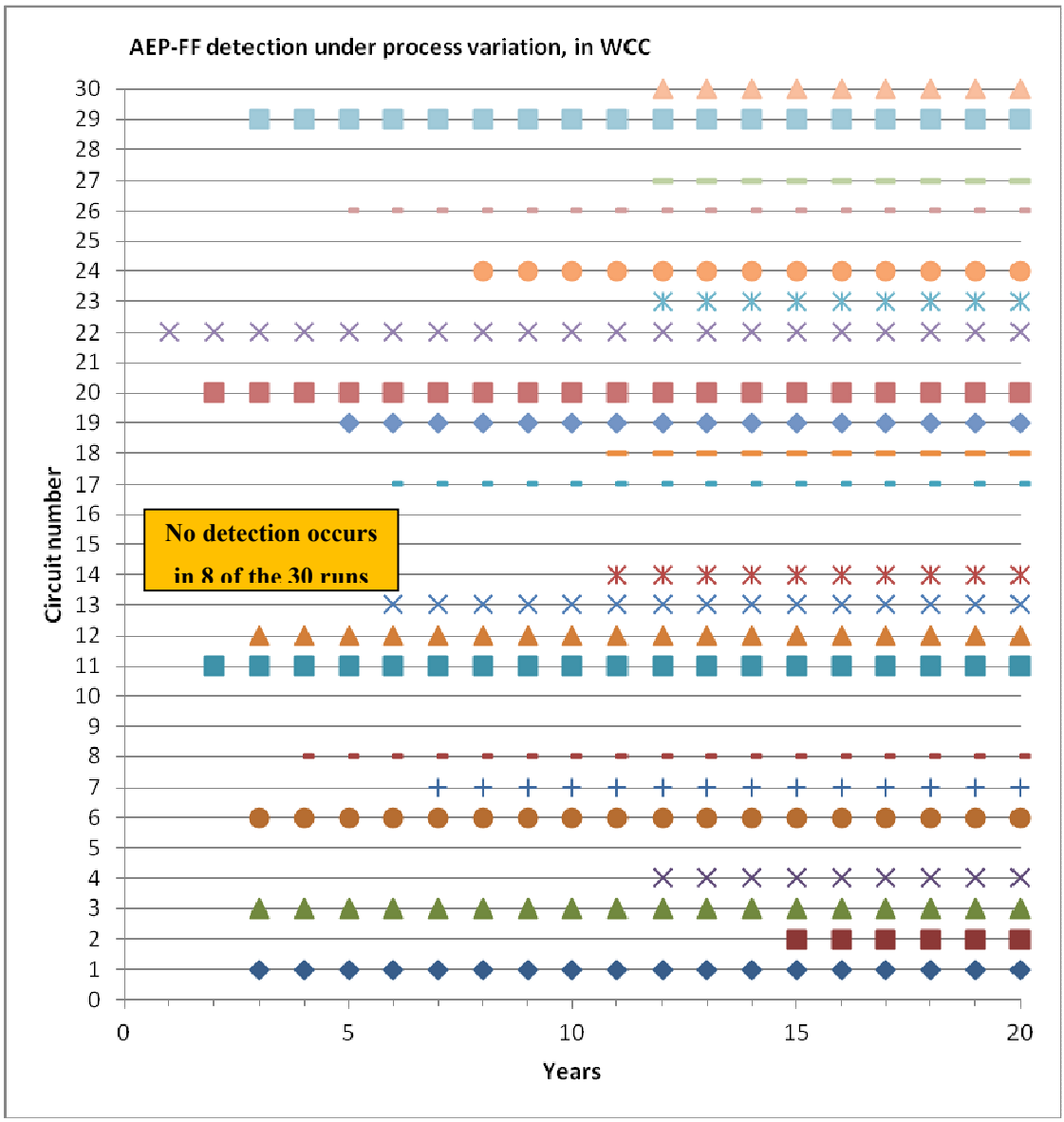


Figure 5.10.AEP-FF guard-band violation detection versus aging, for each MC individual run.

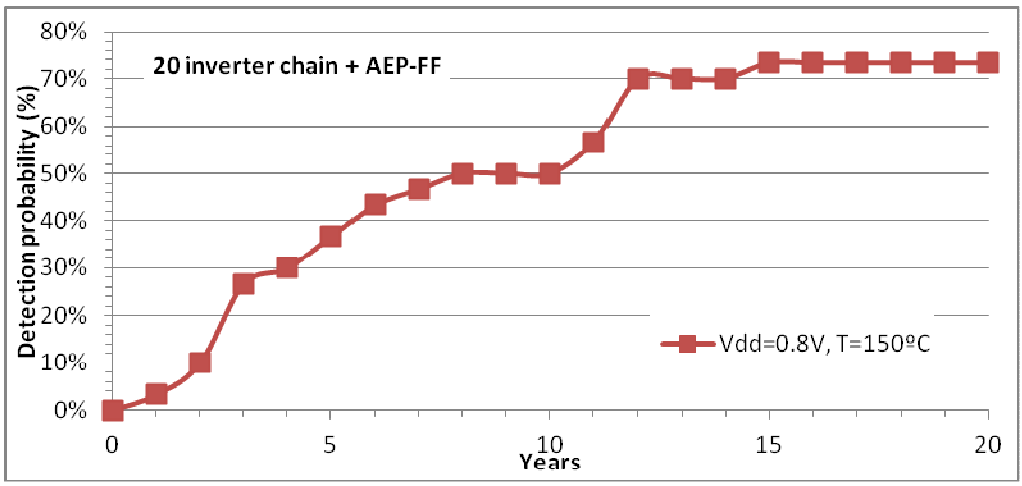


Figure 5.11. Monte-carlo simulation AEP-FF detection probability vs process variation

5.2 PIPELINE MULTIPLIER

A 2-stage, 4-bit pipeline multiplier (PM) was used as the second benchmark circuit (Figure 5.12).

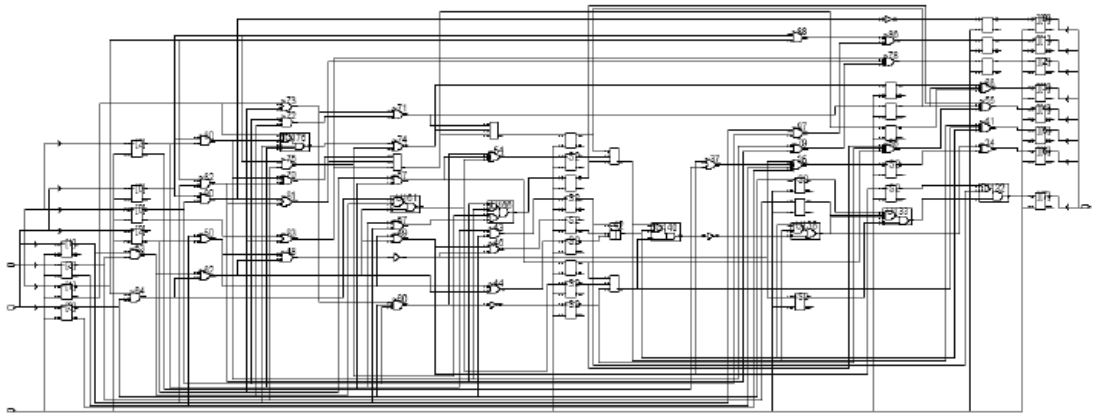


Figure 5.12. 2-stage, 4-bit pipeline multiplier.

The 3 most critical memory elements were identified using Agingcalc, for 12 years NBTI induced aging degradation, with 100Hz stress frequency. The three CME were 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 allow 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. This represents 4.2% path coverage.

The three CME were replaced by three AEP-FF with DE_M, and HSpice simulations were performed with a clock signal period T_{CLK} = 675ps (f_{CLK} = 1.48GHz), where 85% corresponds to the propagation delay of the CP under WCC.

Figure 5.13 depicts the WC propagation time for 3 different paths terminating at CM1, CME2 and CME3. Variations up to 43% are visible for 20 years operation time, due to NBTI induced time degradation (Figure 5.14).

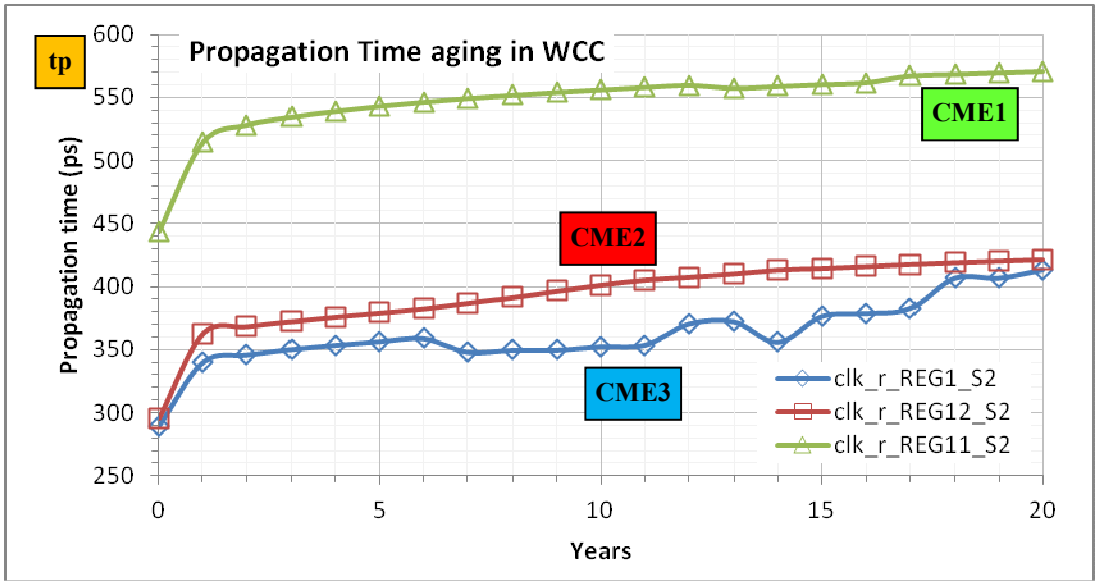


Figure 5.13. Absolute PM propagation time aging, in WCC

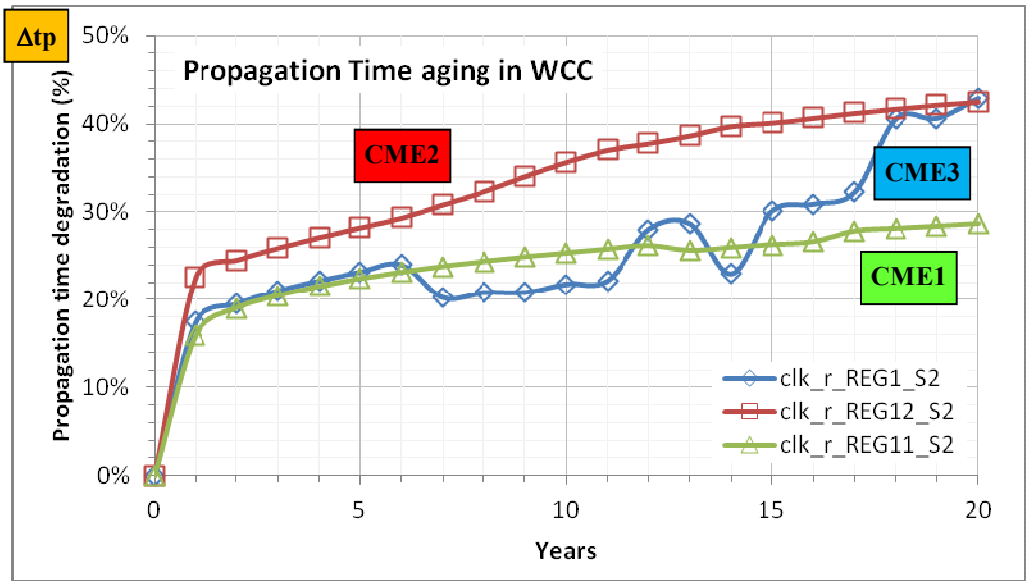


Figure 5.14. Percentage PM propagation time aging, in WCC

5.2.1 CIRCUIT FAILURE PREDICTION

During the 20 years operation time the aging sensors' outputs were monitored, looking for guard-band violation detections. The high delay difference, between CME3 and CM1/CME2, resulted in significant detection period differences between the aging sensor at CME3 and the other two sensors. Detections for CME 3 started at the first operation year, while CME1 started signaling guard-band violations in the 12th year and CME2 in the 19th year (Figure 5.15). Although CME1 and CME2 have similar delay times, the first-detections occur with a 7 year time difference. Two reasons can explain this behavior: first the different paths may have different rise/fall times resulting in different guard-bands (not measured); second, the aging-sensors delay elements will age differently and the delayed signals will arrive at different times to the SC inputs, over the years (Figure 5.16 and Figure 5.17). Comparing Figure 5.15 with Figure 5.17, it is possible to find that a total delay time[9] around 729ps is the requirement to trigger the aging sensors' outputs.

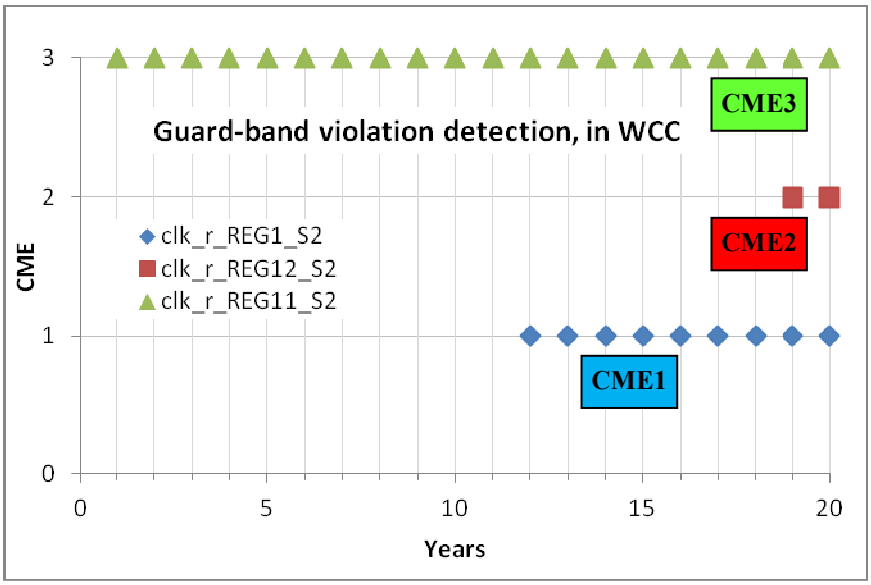


Figure 5.15. Effective detection ranges for 20 years operation time, considering NBTI aging.

[9] Consider the total delay time as the time difference between the clock rising edge and the moment when delayed data signal arrives to the aging sensor's SC input.

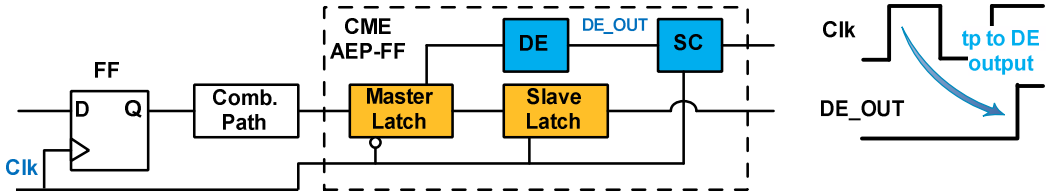


Figure 5.16. PM propagation time to DE output configuration.

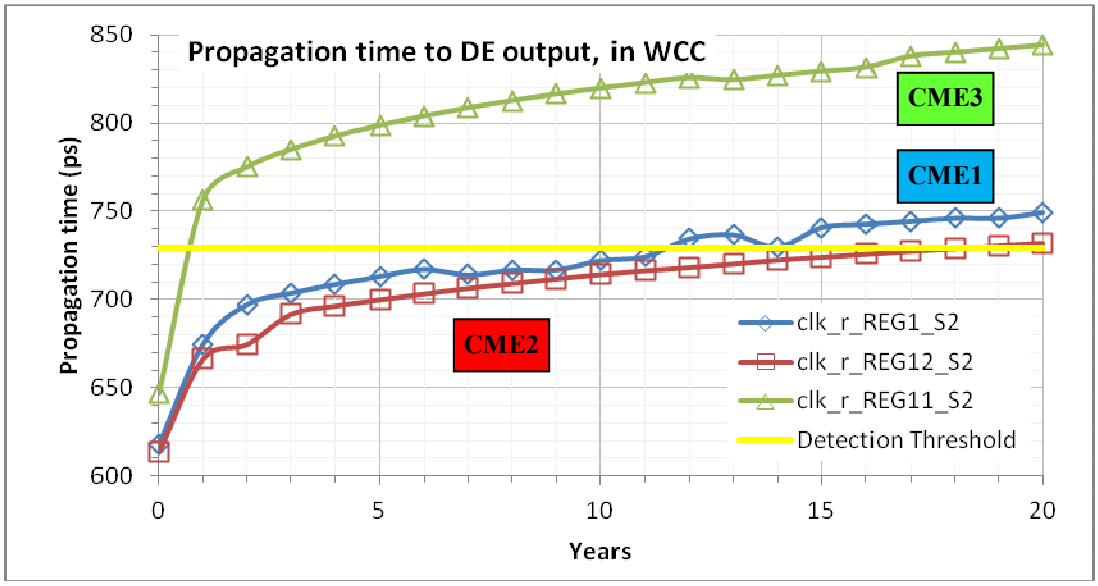


Figure 5.17.Propagation time for the data signal arrive to the SC input (DE output), over 20 years operation in WCC.

5.3 SENSORS COMPARISON

Considering the AEP-FF simulation results, when compared to the previous circuits proposed by Agarwal et. al. [37] and Vazquez et. al. sensors [27], the new aging sensor is globally a better circuit, overcoming many of the other sensors limitations.

On the circuit side, it is integrated in a FF with a binary sensor output to simplify integration with additional monitoring circuitry, like the other two, but uses fewer transistors (23 versus 51 and 44) and doesn't require an output latch.

The monitoring procedure is similar, detecting abnormal data signal delays to predict circuit failure, but the AEP-FF do not load the signal paths, having negligible

to null impact on the circuit's performance (less than 0.2% for the 20 inverters chain circuit).

The guard-band is virtual, defined by a simple delay buffer, requiring no special design or attention to be VTA resilient. This turns guard-band time tuning more simple and fast to be accomplished during design stages. Moreover, the guard-band is not synchronized with the system's clock and do not require an extra guard-band trigger signal to be routed between several aging sensors on the circuit.

A disadvantage is that each AEP-FF needs one DE circuit, which is not critical if we notice that the proposed DE circuits are limited to 4 or 8 transistors.

A particular characteristic and major advantage of the new aging sensor is that the guard-band is enhanced (increases) by VTA variations that degrade the CUT performance, i.e. VTA variations works in favor of AEP-FF failure-prediction. Due to this, the sensor may be active in continuous, assuring that no abnormal delay is missed when a critical path is activated, which is something that is not possible to accomplish with the other two sensors, since they can only be active during short periods of time to avoid aging.

In Table 5.2 are listed all the main parameters for the three aging sensors, highlighting the AEP-FF advantage whenever adequate.

Parameter	**AEP-FF (2011)**	**Agarwal (2007)**	**Vazquez (2010)**
Performance impact	Reduced loading effect	Loading effect in the FF input	
Performance Impact	0% in WCC <0.2% in NC	<1%	-
Area Impact (without the FF)	With DE_L: 19 With DE_M: 23 With DE_H: 23	51 transistors 24 with shared DE	43 transistors 22 with shared DE
$f_{CLK\text{-}min}$ to avoid false	> 16.7MHz, for 65nm technology	> 2MHz, for 90nm technology	-

Parameter	AEP-FF (2011)	Agarwal (2007)	Vazquez (2010)
positives due to charge leakage in SC			
Aging	Enhances guard-band	Degrades guard-band	
VT variations	Enhances guard-band	Degrades guard-band	
Process variations	Nominal values occurs 13.6%, 17.6% and 14% of the time	Nominal value occurs 12% of the time	-
DE	One per sensor	May be shared by several sensors	
Guard-band	Virtual	Synchronous with the clock signal	
Guard-band value	Increases with VTA degradation	Constant	Programmable
Operation	Always on if required	Active from time to time to avoid aging	
Guard-band tuning	Medium (Simple for DE_L and DE_M)	Complex tuning of 3 NAND gates	Simple (2 capacitances)
Output	Binary	Binary	Binary
Concept	Abnormal delay detection		
Latch	Built-in	Additional circuit	

Table 5.2. Aging Sensors comparative table.

6. CONCLUSIONS AND FUTURE WORK

6.1 CONCLUSIONS

This thesis focused on the development of an adaptive error-prediction aging sensor for synchronous digital circuits.

New node technologies experience exponential raise on variability and IC sensitivity to external (and internal) disturbances. This variability seriously affects a digital circuit performance, and its reliable operation. Over time, the circuit ages, which is reflected in longer propagation times in the internal combinational paths, degrading circuit's performance. Between all aging effects, NBTI is referred as the main contributor for the performance loss, in nanoscale technologies using SiO2 dielectric with polysilicon gate devices. Operating conditions like temperature and supply voltage also affects circuit's performance, as high temperatures enhance aging and lower voltages slows down circuit's performance. Moreover, the small device features in nanoscale devices makes process variability a high contributor to the circuit operation degradation. Process, power supply Voltage, Temperature and Aging (PVTA) are four parameters that can influence enormously the performance of an IC and degrade circuit's dependability, as they are operation-dependent and difficult to predict in the design environment. Moreover, dependability during product lifetime requires in-field dependability monitoring.

In this work, attention was focused on these PVTA variability parameters, to know how their effect on circuit's performance degradation can be used to enhance circuit's delay fault prediction capability, when using new performance sensors. The goal was to develop new performance sensors, identified as aging sensors, due to their aging-aware adaptive sensibility. The new sensors predict the occurrence of performance errors (predictive delay-faults) caused by slow cumulative aging degradations, by signalizing late transitions at the data input of key memory cells. The error prediction is made when sensor signalizes the late transitions before the error occurs. The adaptive behavior of the sensor comes from the fact that the sensor's sensibility is

enhanced with aging degradations (or generally, performance degradations) over the sensor itself. Moreover, performance degradation caused by aging effects can also be enhanced by static process variations and/or static/dynamic VT variations.

Circuit failure prediction is growing as a very promising solution and this approach depends on aging sensors to monitor circuit performance loss. Several solutions were already proposed in literature and studied in this work, namely Agarwal et al. work [37] and J. C. Vazquez et. al. work [27]. However, these solutions have several disadvantages, as it was highlighted in this thesis, and the search for the best aging sensor design is what drives research in this area.

The sensor developed in the present work changes the concept and the paradigm associated with the sensor operation, enhancing previous sensors characteristics. In this way the AEP-FF became an innovation.

The first important feature of the AEP-FF is that all the sensor circuitry is installed locally, without the need of additional control signals, or additional synchronizing signals distributed to all the sensors in a circuit.

A second important feature comes from the fact that, although we need to define a guard-band placed at the end of the clock cycle to identify the late transients, the signaling and identification of the late transients is made after the clock cycle's active edge. This is achieved with a new and simple delay element (a common buffer) that delays internally the FF input data transition and creates a virtual guard-band. This operation feature brings additional improvements to the sensor, namely its adaptive sensibility to PVTA variations, and changes the paradigm of sensor design. In the presence of enhanced variability in PVTA degradations, the sensor increases its sensitivity to predict performance errors by increasing locally the virtual guard-band, according with the local PVTA variations. This feature simplifies also sensor design, as the sensor itself does not need to be more robust to PVTA variations than the circuit where it is installed. The guard-band is said to be virtual because there is explicitly no signal defining it and the clock signal is the only operation trigger.

A third important feature relates to sensor's simplicity of usage. It was also demonstrated in this thesis the sensors applicability, being easy to tune and install. The development of three DE circuits simplifies the designer effort in defining a specific guard-band interval for an application, with the added gain that the DE does

not need to be age resilient. The design of a new FF with the sensor functionality included, to directly replace an existing critical memory element, simplifies its installation and applicability to a circuit.

A fourth important feature is the low area overhead they introduce to a circuit. The small number of transistors in the sensor contributes to a small overhead, compared to previous solutions.

The fifth important feature is the negligible performance penalty. Since the aging sensor modules capture the delayed signal on the flip-flop internal nodes, the sensor present low to none performance impact on the paths being monitored.

A sixth important feature is that these local sensors can be always active, increasing circuit's fault-tolerance capability.

All these important features were demonstrated using two benchmark circuits to validate sensor functionality. It was shown that the 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.

As a general and personal conclusion, this work introduced a new electronics expertise field to me, and made a major contribution to my superior education. The opportunity to contribute with a new solution to address technology performance and reliability issues, with potential to become a real industry solution, was indeed a great motivation.

6.2 FUTURE WORK

No research work is completely finished, as many questions normally arise and new topics may be defined, when the work is being pursuit. I do believe the AEP-FF design is a success, but further work need to be done for the design become a real option to the industry, to address reliability and performance issues. This section summarizes possible work as a natural extension of the work developed in this M.Sc. thesis.

A first natural extension of the work is, after the proof of concept, to pursue design automation. A new software tool may be developed to perform automatic sensor insertion in a circuit. The methodology for sensor insertion was already defined in the previously published works ([38, 41]) by the research team. Moreover, the aging prediction analysis and aging-aware critical memory cells identification was also developed in another M.Sc. thesis ([43]). However, circuit automatic reconfiguration by replacing the identified critical FF with AEP-FF is an open task and it can be performed at gate level using a circuit's HDL description.

A second future work topic is related with pre-installed sensor's validation. Circuit simulation and sensor functionality validation requires specific test vectors to activate the critical paths and allow sensor activation. These test vectors are, typically, pairs of vectors, to allow the activations of delay faults. Therefore, after sensor insertion, it is necessary to develop test vectors for the new circuit, not only for circuit simulation during circuit development, but also (eventually) circuit's aging characterization during product's life time. This work is being done in another M.Sc. thesis work, with the definition of a new software tool.

A third future work is the layout design of the new AEP-FF cell and the integration of the new cell in a commercial design flow. The adoption of the new cell by the industry requires a complete cell design and integration in a commercial cell library. With the new cell layout, a complete layout of a CUT (Circuit Under Test) with the aging sensors previously inserted is possible. The next step would be the capacitive extraction and post-layout simulation, for circuit and sensor validation.

The use of commercial library cells by the industry brings an additional problem: the designer may not have access to standard cell layout, which difficults the use of new AEP-FF cells in automatic layout generating design tools. Therefore, a fourth future work can be a new AEP-FF sensor design with standard library cells, to allow easier industrial applicability of the AEP-FF in the commercial design flow.

A fifth future work is silicon validation. The proof of concept for the AEP-FF is only finished when a test chip is produced and real silicon tests were performed. Unfortunately, test chips are very expensive, especially when considering 65nm technologies and below.

A sixth and seventh tasks for future work were already started in reference [41], and is part of another M.Sc. thesis work already started at ISE-UAlg. One task is to

adapt local sensors to be used as global sensors in a circuit. The main advantage of global sensors is that they are less intrusive to the original circuit and easier adopted by the industry. Therefore, using a dummy critical path monitored with a local sensor, by activating periodically (or in every clock period) the path, it is possible to monitor aging degradations. However, these global sensors have disadvantages over the local sensors when local variations differ spatially and in time in the CUT.

The other task is to define a performance and/or power optimization methodology, using local aging sensors, global sensors or both, to define a new aging-aware DVFS methodology. In [41], the use of local and global sensors to tune automatically the power consumption or circuit performance is a promising solution for high variability nanometer technologies (although a more complex solution).

Finally, an eighth task for future work can be identified. Our AEP-FF sensor detects the late transitions, signalizing predictive delay faults. However, the errors do not occur. Hence, it would be of great advantage if, in the presence of an error caused by an abnormal PVTA degradation, or even a Single Event Upset, that the FF could identify the presence of the error and correct it (like previously published Razor II technique [68]). Therefore, additional changes would be necessary over sensor design to account for the new functionality. In fact, one on-going M.Sc. thesis work is trying to implement this new feature, targeting a new sensor with enhanced tolerance to delay faults.

REFERENCES

[1] M. White and Y. Chen, "Scaled CMOS Technology Reliability Users Guide", JPL Publication 08-14 3/08, 2008.

[2] N. Weste and D. Harris, "CMOS VLSI Design: A Circuits and Systems Perspective", Addison Wesley, 2010.

[3] G. Moore, "Cramming More Components Onto Integrated Circuits", Proceedings of the IEEE, vol. 86, pp. 82 –85, January 1998.

[4] Intel, "3-D, 22nm: New Technology Delivers An Unprecedented Combination of Performance and Power Efficiency", http://www.intel.com/content/www/us/en/silicon-innovations/intel-22nm-technology.html, 2012.

[5] M. Bohr and K. Mistry, "Intel™ Revolutionary 22nm Transistor Technology", Intel, May 2011.

[6] G. Shahidi, "Evolution Of CMOS Technology At 32nm And Beyond", in Custom Integrated Circuits Conference, IEEE, pp. 413–416, September 2007.

[7] B. Paul, K. Kang, H. Kufluoglu, M. Ashraful Alam, and K. Roy, "Temporal Performance Degradation Under NBTI: Estimation And Design For Improved Reliability Of Nanoscale Circuits", Vol. 1, pp. 1–6, March 2006.

[8] S. Mitra And M. Agarwal, "Circuit Failure Prediction To Overcome Scaled CMOS Reliability Challenges", IEEE International Test Conference, October 2007.

[9] R. Vattikonda, W. Wang, And Y. Cao, "Modeling And Minimization Of PMOS NBTI Effect For Robust Nanometer Design", Proceedings Of The 43rd Annual Design Automation Conference, pp. 1047–1052, 2006.

[10] Mentor Graphics, "Design For Variability: Managing Design, Process, And Manufacturing Variations In Physical Design", Place And Route Whitepaper, pp. 1-11, 2008.

[11] A. Asenov, S. Kaya, J. Davies, And S. Saini, "Oxide Thickness Variation Induced Threshold Voltage Fluctuations In Decanano MOSFETs: A 3D Density Gradient Simulation Study", Superlattices And Microstructures, Vol. 28, No. 56, pp. 507–515, 2000.

[12] R. Johnson, J. Evans, P. Jacobsen, J. Thompson, And M. Christopher, "The Changing Automotive Environment: High-Temperature Electronics", Electronics Packaging Manufacturing, IEEE Transactions on, Vol. 27, pp. 164–176, July 2004.

[13] R. Oliveira, J. Semião, I. Teixeira, M. Santos, And J. Teixeira, "On-Line BIST For Performance Failure Prediction Under Aging Effects In Automotive Safety-Critical Applications", 12th Latin American Test Workshop (LATW), March 2011.

[14] J. F. L. C. Semião, "Power-Supply And Temperature Based Methodologies To Improve Tolerance And Detection Of Delay Faults In Synchronous Digital

Circuits." Phd Thesis, Universidade Técnica De Lisboa, Instituto Superior Técnico, 2010.

[15] M. Cuviello, S. Dey, X. Bai, And Y. Zhao, "Fault Modeling And Simulation For Crosstalk In System-On-Chip Interconnects", In Computer-Aided Design, Digest Of Technical Papers, IEEE/ACM International Conference On, pp. 297-303, 1999.

[16] W. Chen, S. Gupta, And M. Breuer, "Test Generation In VLSI Circuits For Crosstalk Noise", Proceedings Of Test Conference, pp. 641–650, October 1998.

[17] F. Vargas, "Design And Test On Chip For EMC", Design Test Of Computers, IEEE, Vol. 23, pp. 502–503, June 2006.

[18] S. Zhao And K. Roy, "Estimation Of Switching Noise On Power Supply Lines In Deep Sub-Micron CMOS Circuits", In Thirteenth International Conference On VLSI Design, pp. 168–173, 2000.

[19] H. Chen And L. Wang, "Design For Signal Integrity: The New Paradigm For Deept!Lubmicron VLSI Design", Proceedings Of International Symposium On VLSI Technology, Systems And Applications, pp. 329–333, June 1997.

[20] M. Rodriguez-Irago, J. Andina, F. Vargas, M. Santos, I. Teixeira, And J. Teixeira, "Dynamic Fault Test And Diagnosis In Digital Systems Using Multiple Clock Schemes And Multi-Vdd Test", In 11th IEEE International On-Line Testing Symposium, pp. 281–286, July 2005.

[21] W. Wang, S. Yang, S. Bhardwaj, R. Vattikonda, S. Vrudhula, F. Liu, And Y. Cao, "The Impact Of NBTI On The Performance Of Combinational And Sequential Circuits", pp. 364–369, 2007.

[22] K. Kang, S. Gangwal, S. Park, And K. Roy, "NBTI Induced Performance Degradation In Logic And Memory Circuits: How Effectively Can We Approach A Reliability Solution?", In Asia And South Pacific Design Automation Conference, pp. 726–731, March 2008.

[23] K. Bernstein, K. M. Carrig, C. M. Durham, P. R. Hansen, D. Hogenmiller, E. J. Nowak, And N. J. Rohrer, "High Speed CMOS Design Styles", Kluwer Academic Publishers, 1998.

[24] J. P. C. Teixeira, I. C. Teixeira, M. B. Dos Santos, J. F. L. C. Semião, J. R. Andina, M. Irago, F. Vargas, And L. Piccoli, "Aumento De Tolerância Dinâmica De Circuitos Electrónicos Integrados Digitais A Varições De Tensão De Alimentação E De Temperatura", Portuguese Patent 103436, 2006.

[25] J. Semião, J. Freijedo, J. J. Rodriguez, F. Vargas, M. B. Santos, I. C. Teixeira, And J. P. Teixeira, "Mixed-Signal Design Of Dynamic Delay Buffers To Improve Tolerance To Power Supply And Temperature Variations", 2007.

[26] M. K. J. M. H. S. D. Kim, J. Kim, "System And Method For Monitoring Reliability Of A Digital System", Us Patent 7495519, 2009.

[27] 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", In 28th VLSI Test Symposium (VTS), April 2010.

[28] J. Vazquez, V. Champac, A. Ziesemer, R. Reis, J. Semiaì*f* Ando, I. Teixeira, M. Santos, And J. Teixeira, "Predictive Error Detection By On-Line Aging Monitoring", In IEEE 16th International On-Line Testing Symposium, pp. 9–14, July 2010.

[29] M. Agarwal, V. Balakrishnan, A. Bhuyan, K. Kim, B. Paul, W. Wang, B. Yang, Y. Cao, And S. Mitra, "Optimized Circuit Failure Prediction For Aging:

Practicality And Promise", In IEEE International Test Conference, October 2008.

[30] M. Rajesh, G. Vinod, D. Das, P. Bhatnagar, C. Pithawa, A. Thaduri, And A. Verma, "A Study Of Failure Mechanisms In CMOS And Bjt Ics And Their Effect On Device Reliability", In 2nd International Conference on Reliability, Safety and Hazard, December 2010.

[31] S. N. Wooters, A. C. Cabe, Z. Qi, J. Wang, R. W. Mann, B. H. Calhoun, M. R. Stan, And T. N. Blalock, "Tracking On-Chip Age Using Distributed, Embedded Sensors", In IEEE Transactions On Very Large Scale Integration (VLSI) Systems, Vol. No. 99, pp. 1–12, 2011.

[32] Z. Qi, J. Wang, A. Cabe, S. Wooters, T. Blalock, B. Calhoun, And M. Stan, "Sram-Based NBTI/PBTI Sensor System Design", In 47th ACM/IEEE Design Automation Conference, June 2010.

[33] M. Omana, D. Rossi, N. Bosio, And C. Metra, "Self-Checking Monitor For NBTI Due Degradation", In IEEE 16th International Mixed-Signals, Sensors and Systems Test Workshop, June 2010.

[34] D. K. Schroder And J. A. Babcock, "Negative Bias Temperature Instability: Road To Cross In Deep Submicron Silicon Semiconductor Manufacturing", In Journal Of Applied Physics, Vol. 94, No. 1, pp. 1–18, 2003.

[35] J. Tschanz, J. Kao, S. Narendra, R. Nair, D. Antoniadis, A. Chandrakasan, And V. De, "Adaptive Body Bias For Reducing Impacts Of Die-To-Die And Within-Die Parameter Variations On Microprocessor Frequency And Leakage", In IEEE International Solid-State Circuits Conference, Digest Of Technical Papers, Vol. 1, pp. 422–478 Vol.1, 2002.

[36] S. Das, S. Pant, D. Roberts, S. Lee, D. Blaauw, T. Austin, T. Mudge, And K. Flautner, "A Self-Tuning Dvs Processor Using Delay-Error Detection And Correction", In IEEE Journal of Solid-State Circuits, pp. 792–804, 2006.

[37] M. Agarwal, B. Paul, M. Zhang, And S. Mitra, "Circuit Failure Prediction And Its Application To Transistor Aging", In 25th IEEE VLSI Test Symposium, May 2007.

[38] C. Martins, J. Semiao, J. Vazquez, V. Champac, M. Santos, I. Teixeira, And J. Teixeira, "Adaptive Error-Prediction Flip-Flop For Performance Failure Prediction With Aging Sensors", In 29th IEEE VLSI Test Symposium, pp. 203–208, May 2011.

[39] C. Martins, J. Pachito, J. Semiã£O, I. C. Teixeira, And J. P. Teixeira, "Adaptive Error-Prediction Aging Sensor For On-Line Monitoring Of Performance Errors", In 26th Conference On Design Of Circuits And Integrated Systems, 2011.

[40] J. Pachito, C. Martins, J. Semiao, M. Santos, I. Teixeira, And J. Teixeira, "The Influence Of Clock-Gating On NBTI-Induced Delay Degradation", In IEEE 18th International On-Line Testing Symposium, pp. 61 –66, June 2012.

[41] J. Semiao, J. Pachito, C. Martins, B. Jacinto, J. Vazquez, V. Champac, M. Santos, I. Teixeira, And J. Teixeira, "Aging-Aware Power Or Frequency Tuning With Predictive Fault Detection", In IEEE Design Test Of Computer, Vol. No. 99, p. 1, 2012.

[42] C. V. Martins, J. Semião, I. C. Teixeira, And J. P. Teixeira, "Flip-Flop Com Sensor Para Previsão De Erros De Performance, Para Aplicações Como Sensor Local De Performance Ou De Envelhecimento Em Circuitos Integrados Digitais Síncronos", Portuguese Patent Pending, April 2012.

[43] J. Dos Santos Pachito, "Metodologia Para Prever O Envelhecimento De Circuitos Digitais", Master Of Science Thesis, Universidade Do Algarve - Instituto Superior De Engenharia, 2011.
[44] K. O. Jeppson And C. M. Svensson, "Negative Bias Stress Of Mos Devices At High Electric Fields And Degradation Of Mnos Devices", In Journal Of Applied Physics, Vol. 48, 1977.
[45] B. E. Deal, M. Sklar1, A. S. Grove1, And E. H. Snow1, "Characteristics Of The Surface-State Charge (Qss) Of Thermally Oxidized Silicon", J. Electrochem. Soc., Vol. 114, pp. 266–274, 1967.
[46] A. Islam, H. Kufluoglu, D. Varghese, S. Mahapatra, And M. Alam, "Recent Issues In Negative-Bias Temperature Instability: Initial Degradation, Field Dependence Of Interface Trap Generation, Hole Trapping Effects, And Relaxation", In IEEE Transactions On Electron Devices, Vol. 54, pp. 2143-2154, September 2007.
[47] S. Khan And S. Hamdioui, "Temperature Dependence Of NBTI Induced Delay", In IEEE 16th International On-Line Testing Symposium (IOLTS), pp. 15–20, July 2010.
[48] S. V. Kumar, C. H. Kim, And S. S. Sapatnekar, "An Analytical Model For Negative Bias Temperature Instability", In IEEE/ACM International Conference on Computer-Aided Design, pp. 493–496, November 2006.
[49] M. A. Alam And S. Mahapatra, "A Comprehensive Model Of PMOS NBTI Degradation", Microelectronics Reliability, Vol. 45, No. 1, pp. 71–81, 2005.
[50] Seyab And S. Hamdioui, "NBTI Modeling In The Framework Of Temperature Variation", In Design, Automation & Test in Europe Conference & Exhibition, pp. 283–286, March 2010.
[51] B. L. Sopori, K. M. Jones, X. Deng, R. Matson, M. Al-Jassim, And S. Tsuo, "Hydrogen In Silicon: Diffusion And Defect Passivation", Conference Record of the Twenty Second IEEE Photovoltaic Specialists Conference, October 1991.
[52] H. Kufluoglu, "MOSFET Degradation Due To Negative Bias Temperature Instability (NBTI) And Hot Carrier Injection (HCI) And Its Implications For Reliability-Aware VLSI Design", Phd Thesis, Purdue University, 2007.
[53] H. Dadgour And K. Banerjee, "Aging-Resilient Design Of Pipelined Architectures Using Novel Detection And Correction Circuits", In Design, Automation & Test in Europe Conference & Exhibition, March 2010.
[54] B. Cretu, F. Balestra, And G. Ghibaudo, "Hot Carrier Stress In Deep Submicrometer MOSFET's", In International Electron Devices Meeting, Technical Digest, Vol. 2, October 2001.
[55] S. P. Ching, C. T. Ping, And Y. H. Sun, "Studies Of The Critical LDD Area For HCI Improvement", In IEEE International Conference on Semiconductor Electronics, November 2008.
[56] H. J. Lee And K. K. Kim, "Analysis Of Time Dependent Dielectric Breakdown In Nanoscale CMOS Circuits", In International SoC Design Conference, November 2011.
[57] H. Luo, Y. Wang, J. Velamala, Y. Cao, Y. Xie, And H. Yang, "The Impact Of Correlation Between NBTI And TDDB On The Performance Of Digital Circuits", In IEEE 54th International Midwest Symposium on Circuits and Systems, August 2011.

[58] A. Seth, "Electromigration In Integrated Circuits", "http://www.masamb.com/EM-IITD.pdf", 2009.
[59] J. Lienig, "Introduction To Electromigration-Aware Physical Design", In Proceedings of the 2006 international symposium on Physical design, pp. 39-46, 2006.
[60] K. Bowman, J. Tschanz, C. Wilkerson, S.-L. Lu, T. Karnik, V. De, And S. Borkar, "Circuit Techniques For Dynamic Variation Tolerance", In 46th ACM/IEEE Design Automation Conference, July 2009.
[61] A. Jakubowski And L. Å Ukasiak, "CMOS Evolution Development Limits", Materials Science, Vol. 26, 2008.
[62] S. Doh, H. Jung, And Y. Kim, "Improvement Of NBTI And Electrical Characteristics By Ozone Pre-Treatment And PBTI Issues In Hfalo(N) High-K Gate Dielectrics", In IEEE International Electron Devices Meeting, Technical Digest, pp. 38.7.1–38.7.4, December 2003.
[63] A. Scarpa, D. Ward, J. Dubois, L. Van Marwijk, S. Gausepohl, R. Campos, K. Y. Sim, A. Cacciato, R. Kho, And M. Bolt, "Negative-Bias Temperature Instability Cure By Process Optimization", In IEEE Transactions On Electron Devices, Vol. 53, pp. 1331–1339, June 2006.
[64] J. Hicks, D. Bergstrom, And H. M., "45nm Transistor Reliability", Intel Technology Journal, Vol. 12, P. 16, 2008.
[65] W. Kuo And T. Kim, "An Overview Of Manufacturing Yield And Reliability Modeling For Semiconductor Products", Proceedings Of The IEEE, Vol. 87, pp. 1329 –1344, August 1999.
[66] A. Vassighi, O. Semenov, M. Sachdev, A. Keshavarzi, And C. Hawkins, "CMOS IC Technology Scaling And Its Impact On Burn-In", In IEEE Transactions On Device And Materials Reliability, Vol. 4, pp. 208–221, June 2004.
[67] Z. Amini-Shehsdeh And A. Nabavi, "A Novel Sensor For Prediction Of Aging Failure", In Third International Conference on Computational Intelligence, Modelling and Simulation, September 2011.
[68] D. Blaauw, S. Kalaiselvan, K. Lai, W.-H. Ma, S. Pant, C. Tokunaga, S. Das, And D. Bull, "Razor Ii: In Situ Error Detection And Correction For Pvt And Ser Tolerance", In IEEE International Solid-State Circuits Conference, Digest Of Technical Papers, pp. 400 –622, February 2008.
[69] M. Hashimoto, "Run-Time Adaptive Performance Compensation Using On-Chip Sensors", In 16th Asia and South Pacific Design Automation Conference, January 2011.
[70] S. Wang, M. Tehranipoor, And L. Winemberg, "In-Field Aging Measurement And Calibration For Power-Performance Optimization", In 48th ACM/EDAC/IEEE Design Automation Conference, June 2011.
[71] G. Yan, Y. Han, And X. Li, "Revivenet: A Self-Adaptive Architecture For Improving Lifetime Reliability Via Localized Timing Adaptation", In IEEE Transactions On Computers, Vol. 60, pp. 1219–1232, September 2011.
[72] M. Hashimoto And H. Fuketa, "Adaptive Performance Compensation With On-Chip Variation Monitoring", In IEEE 54th International Midwest Symposium on Circuits and Systems, August 2011.
[73] M. Civilini, "Reliability And Field Aging Time Using Temperature Sensors", In Fourth International Conference on Sensor Technologies and Applications, July 2010.

[74] A. Cabe, Z. Qi, S. Wooters, T. Blalock, And M. Stan, “Small Embeddable NBTI Sensors (SENS) For Tracking On-Chip Performance Decay”, In Quality of Electronic Design, ISQED, March 2009.

[75] J. Keane, T. Kim, And C. Kim, “An On-Chip NBTI Sensor For Measuring PMOS Threshold Voltage Degradation”, In Proceedings of the International Symposium on Low Power Electronics and Design , August 2007.

[76] K. Kim, W. Wang, And K. Choi, “On-Chip Aging Sensor Circuits For Reliable Nanometer MOSFET Digital Circuits”, In IEEE Transactions On Circuits And Systems II: Express Briefs, Vol. 57, pp. 798 –802, October 2010.

[77] K. Ramakrishnan, X. Wu, N. Vijaykrishnan, And Y. Xie, “Comparative Analysis Of NBTI Effects On Low Power And High Performance Flip-Flops”, In IEEE International Conference on Computer Design, October 2008.

[78] “Http://Coe.Berkeley.Edu/About/History-And-Traditions/1972-The-Release-Of-Spice.Html”, 2012.

[79] P. Park, J. John, K. Klein, J. Teplik, J. Caravella, J. Whitfield, K. Papworth, And S. Cheng, “Reversal Of Temperature Dependence Of Integrated Circuits Operating At Very Low Voltages”, In International Electron Devices Meeting, pp. 71 –74, December 1995.

[80] A. Bellaouar, A. Fridi, M. Elmasry, And K. Itoh, “Supply Voltage Scaling For Temperature Insensitive CMOS Circuit Operation”, In IEEE Transactions On Circuits And Systems II: Analog And Digital Signal Processing, Vol. 45, pp. 415 –417, March 1998.

[81] X. Fu, T. Li, And J. Fortes, “NBTI Tolerant Microarchitecture Design In The Presence Of Process Variation”, In 41st IEEE/ACM International Symposium On Microarchitecture, pp. 399 –410, November 2008.

[82] J. Abella, X. Vera, And A. Gonzalez, “Penelope: The NBTI-Aware Processor”, In Proc. 40th IEEE/ACM International Symposium On Microarchitecture, pp. 85–96, December 2007.

[83] J. Semião Et Al., “Delay-Fault Tolerance To Power Supply Voltage Disturbances Analysis In Nanometer Technologies”, Proc. IEEE Int. On-Line Testing Symp. (IOLTS), pp. 223-228, 2009.

[84] 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.

[85] M.A. Alam, H. Kufluoglu, D. Varghese, And S. Mahapatra. “A Comprehensive Model For PMOS NBTI Degradation: Recent Progress”. Microelectronics Reliability, pp. 853–862, 2007.

APPENDIX

PATENTE

MICROELECTRÓNICA

CÓDIGO: 1838196

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

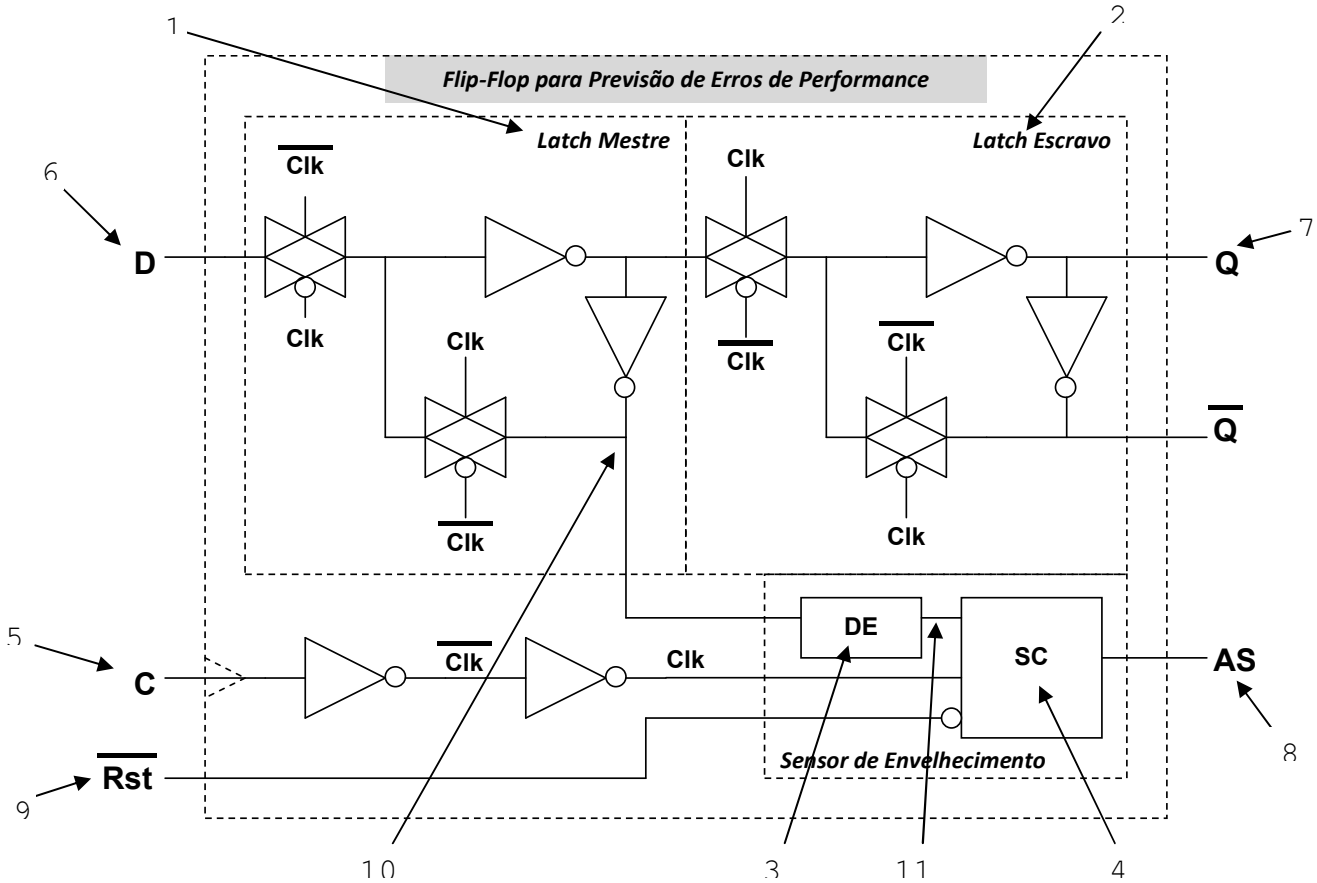
RESUMO

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.

A presente invenção refere-se a um Flip-Flop com sensor para previsão de erros de performance, para aplicação como sensor de performance ou de envelhecimento em circuitos integrados digitais síncronos desenvolvidos em nanotecnologias CMOS. A presente invenção é composta por um Flip-flop mestre-escravo (1, 2), um elemento de atraso (3) e um elemento detector de estabilidade (4). O principal objectivo é detectar transições de dados tardias na entrada do flip-flop (6), constituindo assim um sensor de performance preditivo para circuitos digitais síncronos, que pode prever erros provocados pelo envelhecimento lento dos transístores que compõem o circuito integrado. A principal vantagem é o sensor tornar-se mais sensível às transições tardias na entrada do flip-flop (6), à medida que o próprio sensor envelhece, ou que a tensão de alimentação diminui, ou que a temperatura aumente. O sensor adapta-se, aumentando a sua sensibilidade, à medida que as condições de funcionamento do circuito fazem aumentar a probabilidade de ocorrência de erros de performance.

Faro, 17 de Abril de 2012

FIGURA PARA PUBLICAÇÃO



DESCRIÇÃO

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.

Área técnica de invenção

Os circuitos electrónicos actuais são constituídos, maioritariamente, por transístores MOSFET (Metal Oxide Semiconductor Field Effect Transistor). Em particular, os circuitos digitais actuais utilizam de uma forma generalizada este tipo de transístores, nomeadamente os transístores do tipo PMOSFET (canal do tipo P) e NMOSFET, (canal do tipo N) formando a lógica CMOS muito utilizada hoje em dia para implementar funções lógicas para solucionar problemas complexos. Os circuitos digitais mais comuns são os circuitos digitais síncronos e, neste tipo de circuitos, um sinal de relógio é utilizado para coordenar e sincronizar o funcionamento de todo o circuito, fazendo com que a transmissão de sinais entre os elementos de memória aconteça em momentos pré-definidos. Isto permite prever e coordenar mais facilmente a operação de circuitos complexos, com múltiplas ligações e diversos dados trocados entre os diferentes blocos constituintes do circuito. O sincronismo confere assim uma previsibilidade temporal para a realização de tarefas e facilita a comunicação entre diferentes sistemas. Por outro lado, a falha deste sincronismo confere uma vulnerabilidade maior deste tipo de sistemas à ocorrência de erros de desempenho.

Nos últimos anos os circuitos digitais CMOS têm evoluído continuamente para circuitos cada vez mais pequenos, sendo que actualmente as tecnologias nanométricas permitem construir circuitos em que o comprimento do canal dos transístores é inferior a 100nm. Esta redução crescente do tamanho do canal tem permitido aumentar a performance dos circuitos digitais, verificando-se o aumento da frequência de trabalho do sinal de relógio dos circuitos digitais síncronos.

No entanto, com as tecnologias nanométricas surgem também novos problemas e desafios no que diz respeito fiabilidade dos circuitos CMOS e, neste aspecto, destaca-se a variabilidade como um factor que afecta negativamente essas propriedades. A variabilidade representa todas as incertezas, inconsistências, desequilíbrios e desvios associados a um determinado parâmetro ou característica de um circuito, e provoca incertezas no seu funcionamento, em particular na sua performance. De uma forma geral, tem como causas principais: a variabilidade no processo de fabrico (variações na tensão limiar de condução dos transístores, Vth, comprimento, L, e largura, W, do canal, entre outros); a variabilidade operacional (múltiplos modos de concepção, controlo dinâmico da tensão de alimentação e frequência, entre outros); a variabilidade nas condições de operação do circuito (variações na tensão de alimentação e temperatura de funcionamento, diafonia, quedas de tensão nos condutores, entre outros); e a variabilidade a longo prazo (efeitos do envelhecimento dos circuitos, em especial por NBTI – Negative Bias Temperature Instability, radiação, entre outros). Assim, devido à variabilidade, é de esperar uma menor confiança

e fiabilidade nos circuitos quando se utilizam as escalas nanométricas.

O envelhecimento dos circuitos é um processo de degradação lento, a longo prazo, que pode afectar a performance dos circuitos integrados durante a sua vida útil, e pode também activar defeitos que ficaram latentes desde o fabrico. Uma das causas de envelhecimento com efeito dominante sobre a redução de performance a longo prazo nas tecnologias nanométricas é denominada de NBTI (Negative Bias Temperature Instability) e provoca o aumento de |Vth| dos transístores PMOS ao longo do tempo. Variações de |Vth| afectam a velocidade com que os transístores podem comutar de estado, o que por sua vez torna os circuitos mais lentos.

Num circuito digital síncrono, as variações de desempenho provocadas por variações do processo de fabrico, da tensão de alimentação, da temperatura e do envelhecimento vão, em ultima instância, deteriorar os tempos de propagação entre elementos de memória e provocar atrasos na transmissão de sinais no circuito integrado. O aumento da temperatura, a diminuição da tensão de alimentação de um circuito e o aumento de |Vth|, aumentam o tempo de propagação dos caminhos. Estas variações temporais podem resultar em dados corrompidos ou falhas no funcionamento do circuito síncrono, porque os sinais deixam de chegar dentro do tempo previsto para serem processados correctamente, isto é, deixa de haver sincronismo entre o relógio do sistema e a propagação dos sinais nos caminhos combinatórios.

A sensibilidade das tecnologias nanométricas à variabilidade obriga a que os projectistas procurem por novas soluções para reduzir os seus efeitos e tornar estas tecnologias fiáveis e aplicáveis em todo o universo

da electrónica. Em sistemas não tolerantes a falhas (ex.: sistemas de segurança e alarme, electrónica automóvel, aparelhos médicos, equipamentos aeroespaciais, entre outros) é obrigatório utilizar circuitos que sejam totalmente fiáveis. Actualmente isso é possível de realizar à custa da perda significativa de performance e/ou área. Uma solução típica utilizada nos circuitos síncronos é aplicar margens de segurança temporais (ou margens de erro) nos caminhos mais longos de um circuito integrado, para absorver as variações temporais induzidas por variações de processo de fabrico, tensão, temperatura ou envelhecimento. As margens a aplicar são definidas na fase de desenho e para o seu cálculo é considerado o pior caso possível de funcionamento do circuito durante o seu tempo de vida útil. No entanto, esta metodologia é demasiado pessimista, pois está-se a limitar a frequência máxima de operação do circuito em todas as condições, não sendo possível explorar ao máximo o potencial da tecnologia. Além disto, podem sempre ocorrer erros se existirem variações anómalas do processo de fabrico, da tensão de alimentação ou da temperatura, acima das previsões.

Para se conseguir potenciar ao máximo a tecnologia, novos métodos têm sido sugeridos e utilizados. Um dos métodos usa 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.

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. No entanto, embora estes métodos possam corrigir alguns erros verificados nos elementos de memória síncronos (como os flip-flop), não podem ser utilizados em todos os circuitos síncronos, pois a correcção de erros pode ser uma tarefa complicada e porque não impedem a ocorrência das falhas (que podem ser críticas em algumas aplicações).

A nova solução envolve a previsão de erros de performance para possibilitar a activação de mecanismos que impeçam a sua ocorrência. 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. Por exemplo, quando associado a um sistema de controlo, a detecção de um potencial erro pode desencadear um processo para diminuir a frequência do sinal de relógio do sistema, ou enviar um sinal de alarme ao utilizador do equipamento onde o circuito integrado está inserido. Esta metodologia recorre a sensores de envelhecimento e sensores de performance, para medir a deterioração do circuito ou do seu desempenho temporal e poder gerar os sinais de alarme. A invenção apresentada neste documento enquadra-se na família dos sensores de performance e sensores de envelhecimento.

Estado da técnica

Devido à grande importância dos efeitos da NBTI na performance dos circuitos digitais, este fenómeno tem sido estudado exaustivamente pela comunidade científica [3][4] e actualmente é possível encontrar modelos que permitem simular o seu funcionamento [4][5][6][16][17]. Para combater este fenómeno, foram desenvolvidas soluções para aumentar a tolerância à degradação de |Vth| [2][6], e apresentados estudos sugerindo técnicas de desenho que limitam a degradação provocada pelo efeito NBTI [3].

Adicionalmente, foram também apresentadas várias topologias para sensores de envelhecimento capazes de, globalmente, detectar num circuito a perda de performance devido ao envelhecimento [7][8][9][10]. Porém 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).

Posteriormente, M. Agarwal et al. [11] propuseram uma nova abordagem ao problema centrada na previsão de erros, com o objectivo final 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), e que limita significativamente a sua performance. Como solução, apresenta um sensor de envelhecimento 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. O sensor é

composto por um elemento de atraso (DE), um detector de estabilidade (SC) e um Latch. O elemento de atraso está sincronizado com o sinal de relógio do sistema e define uma margem temporal de segurança (denominada como Banda de Guarda) dentro da qual o sinal de dados na entrada do flip-flop é monitorizado e não pode variar, caso contrário é sinalizada uma transição tardia que pode estar perto de originar erros de performance; o detector de estabilidade, SC, está ligado na entrada de dados do Flip-flop e faz exactamente a sinalização da transição tardia, activando a sua saída se ocorrer alguma variação do sinal dentro do período definido pelo elemento de atraso; o latch garante que o sinal de saída do detector de estabilidade permanece activo até ordem em contrário. O tempo definido pelo elemento de atraso é sempre programado de forma a garantir que a perda de performance é detectada antes de haver violação do sincronismo no flip-flop.

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 [12][13][14]. 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 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 já existentes 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 (PVT), 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.

Outro tipo de sensores em flip-flops foram ainda definidos em [1][18][19] e [20]. Estes sensores não são

sensores preditivos mas correctivos, ou seja, detectam a degradação de performance pela captura dos dados depois do sinal de sincronismo de relógio dos flip-flops ser activado, e não pela captura de dados tardia mas ainda síncrona com o relógio pelo flip-flop. Para além disso, possuem um latch adicional que permite não só detectar que a primeira captura realizada no flip-flop foi incorrecta, mas também corrigi-la, pois o latch adicional captura o valor correcto dos dados que chegam atrasados. Porém, 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.

A presente invenção enquadra-se nos sensores que prevêem a degradação de performance localmente onde o erro de sincronismo ocorre, ou seja, nos flip-flops. Assim, a presente invenção consiste num Flip-Flop para prever erros de performance em sensores de envelhecimento, para aplicação em circuitos integrados digitais síncronos desenvolvidos em nanotecnologias CMOS. Para além disso, a presente invenção permite resolver muitos dos problemas verificados nas soluções de sensores de performance locais preditivos já existentes. 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.

Sumário da invenção

A presente invenção refere-se a um Flip-Flop para previsão de erros de performance em sensores de envelhecimento, para aplicação em circuitos integrados digitais síncronos desenvolvidos em nanotecnologias CMOS. Permite monitorizar a degradação da performance nos circuitos digitais síncronos, provocada por variações no processo de fabrico, por variações na tensão de alimentação, na temperatura, ou pela degradação do funcionamento dos circuitos provocada pelo seu envelhecimento (nomeadamente as degradações provocadas pelo efeito NBTI).

A presente invenção integra as funcionalidades de flip-flop e de sensor de envelhecimento. Integrado no flip-flop, o sensor de envelhecimento define uma banda de guarda (intervalo de tempo para observação) virtual, independente do sinal de relógio, dentro da qual irá detectar variações no sinal de dados que está a monitorizar. A banda de guarda é identificada como virtual, por não existir nenhum sinal a defini-la explicitamente. A banda de guarda não é programável e é definida inicialmente durante o desenho do circuito, após se terem identificado e caracterizado os vários caminhos críticos a monitorizar. Embora seja inicialmente constante, a banda de guarda adapta-se com as condições de funcionamento do circuito, aumentando quando há degradação dos parâmetros que afectam a performance do circuito. O aumento de temperatura, a diminuição na tensão de alimentação do circuito e o aumento da degradação provocada pelo envelhecimento (nomeadamente pelo aumento de Vthp, devido à NBTI), todos eles,

provocam o aumento da banda de guarda, mudando a sensibilidade do sensor ao longo da vida útil do circuito, e aumentando essa sensibilidade quando a degradação do circuito é maior. A presente invenção pode assim monitorizar constantemente os caminhos críticos, sem haver preocupação com a eventual degradação do sensor e consequente diminuição da sua sensibilidade.

Quando ocorre uma variação tardia (mas dentro do período de relógio) na entrada de dados do flip-flop, ou seja, dentro da banda de guarda, a saída do sensor fica activa para sinalizar essa ocorrência, e mantém-se nesse estado até ser dado um comando para repor a sua saída a zero.

Por outro lado, a utilização da presente invenção requere a substituição de um ou mais flip-flops originais pelos flip-flops definidos na presente invenção, com a funcionalidade de sensor de performance incluída no flip-flop. A inserção do elemento de atraso e do detector de estabilidade no circuito de um flip-flop, não provoca (ou provoca marginalmente) a degradação de performance do circuito original, uma vez que as ligações internas dos diferentes blocos são feitas fora do caminho de dados principal no flip-flop.

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 do princípio de operação da invenção.

– fig.3, fig.4, e fig.5 são três representações de arquitecturas possíveis para o elemento de atraso (DE).

– fig.6 é uma representação da arquitectura do detector de estabilidade (SC).

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.

A presente invenção (fig.1) é composta por um flip-flop genérico (como descrito em [15]) (e composto por dois Latches, Latch Mestre (1) e Latch Escravo (2)), um elemento de atraso (3), um detector de estabilidade (4), uma entrada de relógio (5), uma entrada de dados (6), uma saída de dados (7), a saída do sensor de performance e de envelhecimento (8) e uma entrada de reinicialização do sensor (9).

A presente invenção (fig.1) funciona como um flip-flop do tipo D activo no flanco ascendente do relógio, que é composto por dois Latches, Latch Mestre (1) e Latch Escravo (2), uma entrada C (5) para o sinal de relógio, uma entrada D (6) para o sinal de dados, e uma saída principal de dados Q (7). Durante o meio-ciclo negativo

do sinal de relógio, o sinal (10) do Latch Mestre, captura continuamente o sinal da entrada D (6), e o Latch Escravo (2) isola a saída Q (7) do sinal de dados presente no Latch Mestre (2), mantendo o nível lógico em Q (7) inalterado. Na transição do semi-ciclo negativo para o semi-ciclo positivo, do sinal de relógio, o Latch Mestre (1) isola a sua saída (10) da entrada D (6), mantendo o sinal em (10) constante, e o Latch Escravo (2) actualiza a sua saída (7) com o valor presente no sinal de dados do Latch Mestre (2) (igual a (10)). Durante o restante semi-ciclo positivo, o Latch Escravo continua a capturar o sinal de dados do Latch Mestre, igual ao sinal (10), e o valor em Q (7) fica constante e independente das variações do sinal na entrada D (6).

Para implementar a funcionalidade de sensor de performance e de envelhecimento na presente invenção (fig.1), é adicionado ao flip-flop um elemento de atraso (3), um detector de estabilidade (4), uma saída AS (8) e uma entrada /Rst (9). O elemento de atraso (3) vai atrasar sempre o sinal que se encontra no nó (10) e aplicar esse sinal na entrada (11) do detector de estabilidade (4). Durante o semi-ciclo positivo do sinal de relógio, o detector de estabilidade vai monitorizar o nó (11) e activar a saída AS (8), colocando-a a '1' lógico no caso de o sinal no nó (11) mudar de estado. Depois de activa, a saída AS (8) mantém-se indefinidamente nesse estado até que seja aplicado um '0' lógico na entrada de /Rst (9), quando o sinal de relógio se encontra no semi-ciclo negativo. Nessa altura a saída AS (8) retoma o valor lógico '0'. Para haver detecção a entrada /Rst (9) tem de estar a '1' lógico.

O princípio de funcionamento do sensor (fig.2) implica atrasar o sinal de dados aplicado na entrada D (6) por um

tempo igual à banda de guarda que se pretende implementar (12), durante o semi-ciclo negativo do sinal de relógio. Se houver degradação significativa do caminho crítico (13), qualquer mudança no nível lógico do sinal vai chegar atrasada e dentro do intervalo de tempo definido pela banda de guarda virtual (14). Como este sinal é posteriormente atrasado (15) pelo elemento de atraso (3), essa variação vai reflectir-se no nó (11) no semi-ciclo positivo do sinal de relógio (15,16), e vai ser detectada pelo detector de estabilidade (4), que por sua vez vai sinalizar uma violação na banda de guarda através da saída AS (8). Quando não há degradação do caminho crítico, as variações de sinal vão ocorrer na entrada D (6) antes do intervalo de tempo definido pela banda de guarda e chegam ao nó (11) antes de se iniciar o semi-ciclo positivo do sinal de relógio.

Os elementos do flip-flop, incluindo o Latch Mestre (1) e Latch Escravo (2), são implementados com portas de transmissão e inversores (fig.1).

O elemento de atraso (3) é implementado por um buffer, cuja arquitectura pode ser escolhida de acordo com a capacidade de atraso pretendida, sem necessidade de ser resiliente aos efeitos provocados por variações de processo de fabrico, de tensão de alimentação, de temperatura ou por degradação provocada pelo envelhecimento. Nas figuras são apresentadas três possíveis soluções para o elemento de atraso (3), em que o primeiro circuito (fig. 3) apresenta os maiores tempos de atraso, seguido do segundo circuito (fig.4), sendo o terceiro circuito (fig.5) o que apesenta os tempos de atraso menores (se todos forem desenhados com transístores iguais). Estes quatro circuitos permitem definir tempos de atraso muito pequenos, adequados às

tecnologias nanométricas, a diversas frequências de funcionamento, a diversas sensibilidades para o sensor e a diferentes margens de segurança para o relógio do circuito.

O Detector de estabilidade (fig. 6) é implementado com lógica CMOS dinâmica e inclui lógica para retenção do sinal activo na saída AS (19). Durante o semi-ciclo negativo do sinal de relógio (17), considerando que a saída AS (19) está num nível lógico baixo, os nós X (32) e Y (33) ficam ambos a '1', através dos transístores M6 (26), M7 (27) e M9 (29), e a saída AS (19) mantém-se a '0'. Quando o sinal de relógio (17) comuta para o semi-ciclo positivo, os transístores M3 (23) e M4 (24) estão desligados e, consoante o nível lógico existente no sinal de entrada (18), então M1 (21) e M2 (22), ou M5 (25), ficam ligados e um dos nós, X (32) ou Y (33), comuta para o nível lógico baixo, forçando AS (19) a tomar o valor lógico '1'. A partir deste momento, o transístor M9 (29) fica desligado e deixa de ser possível colocar ambos os nós X (32) e Y (33) no nível lógico '1' no semi-ciclo negativo do sinal de relógio (17), excepto se o sinal na entrada de /Rst (20) estiver num nível lógico baixo. Para reforçar a estabilidade da saída AS (19), os transístores M3 (23) e M4 (24) forçam os nós X (32) e Y (33) a ficarem ambos a '0', comandados pelo sinal alto em AS (19). Para colocar a saída AS (19) novamente no nível lógico baixo, é necessário aplicar um '0' lógico na entrada /Rst (20) durante, pelo menos, um dos semi-ciclos negativos do sinal de relógio (17). A lógica de retenção é implementada pelos transístores M3 (23), M4 (24) e M9 (29).

Esta topologia pode ser expandida e aplicada a outros tipos de flip-flop, e o elemento de atraso pode ser construído com qualquer outro circuito de atraso, cujas características de atraso aumentem com as variações provocadas por degradações na tensão de alimentação, aumento da temperatura ou degradação da condução provocada por envelhecimento dos transístores.

Faro, 17 de Abril de 2012

Referências

[1] 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, nº 4, pp. 792-804, April 2006.

[2] 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.

[3] 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.

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

[5] Weping Wang et. al. “The impact of NBTI on the performance of combinational and sequential circuits”, Proceedings of the 44th annual Design Automation Conference (DAC '07), ACM New York, NY, USA, 2007

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

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

[15] R. J. Baker, “CMOS Circuit Design, Layout, and Simulation”, 2nd Ed., IEEE Press, Wilew-Interscience, 2005.

[16] 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.

[17] 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.

[18] 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.

[19] 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, nº 6, pp 10-20, November 2004.

[20] 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.

Faro, 17 de Abril de 2012

REIVINDICAÇÕES

1 – 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, caracterizado por um flip-flop mestre-escravo, constituído por um Latch Mestre (1) e por um Latch Escravo (2), e por um sensor de performance, constituído por um elemento de atraso (DE) (3), que está ligado internamente no Latch Mestre (1) no seu caminho de realimentação (10), e por um detector de estabilidade (4), que está ligado à saída do elemento de atraso (3).

2 – Flip-flop com sensor de previsão de erros de performance, de acordo com a reivindicação nº 1, caracterizado por ter os portos de entrada e saída comuns de um flip-flop, por ter um porto de entrada para inicialização do sensor, designado por /Rst (9), e por ter um porto de saída para sinalização da previsão de erro pelo sensor, designado por AS (8).

3 – Flip-flop com sensor de previsão de erros de performance, de acordo com a reivindicação nº 1, caracterizado por integrar um elemento de atraso (3), constituído por inversores lógicos, que introduz um atraso nas transições de nível que ocorrem na sua entrada (10), devido à propagação dessas transições pelos inversores lógicos.

4 – Flip-flop com sensor de previsão de erros de performance, de acordo com a reivindicação nº 1, caracterizado por integrar um detector de estabilidade (4) que inclui lógica para retenção do sinal de saída AS

(8) em modo activo implementada por uma realimentação interna (e sem o recurso a uma célula de memória adicional), e que inclui lógica para inicialização do sensor e desactivação do sinal de saída AS (8) controlada por um sinal de entrada /Rst (9) activo no nível baixo.

5 – Flip-flop com sensor de previsão de erros de performance, de acordo com a reivindicação nº 1, caracterizada por ter características de atraso iguais, ou marginalmente iguais, a um flip-flop que não tenha o sensor de performance.

6 – Processo de funcionamento do sensor de performance, utilizado com flip-flops para detectar a degradação de performance provocada pelo envelhecimento dos circuitos digitais síncronos, constituído por um elemento de atraso (3) e por um detector de estabilidade (4), caracterizado por monitorizar as transições de dados tardias, na entrada de dados do Flip-flop (6), e por sinalizar a ocorrência destas transições tardias, na saída AS (8), durante a primeira metade do período de relógio seguinte.

7 – Processo de funcionamento do sensor de performance, de acordo com a reivindicação nº 6, caracterizado por prever e a ocorrência de erros de performance provocados por degradações lentas na performance de um circuito, degradações essas causadas pelo envelhecimento dos circuitos (degradação lenta e cumulativa), e/ou pela variação lenta na temperatura de operação dos circuitos, e/ ou por variações lentas na tensão de alimentação.

8 – Processo de funcionamento do sensor de performance, de acordo com a reivindicação nº 6, caracterizado por

aumentar a sua capacidade para prever as variações tardias do sinal de entrada de dados (6) com o aumento da degradação provocada pelo envelhecimento dos circuitos, e/ou com o aumento da temperatura de operação, e/ou com a diminuição da tensão de alimentação.

9 – Processo de funcionamento do sensor de performance, de acordo com a reivindicação nº 6, caracterizado por sinalizar variações do nível lógico na entrada do detector de estabilidade (11), que ocorram durante o semi-ciclo positivo do sinal de relógio (5), para o caso de um Flip-flop com activação pelo flanco ascendente do relógio, ou que ocorram durante o semi-ciclo negativo do sinal de relógio, para o caso de um Flip-flop com activação pelo flanco descendente do relógio, colocando a saída do detector de estabilidade, AS (8), no nível lógico '1' quando essas condições se verificam e no nível lógico '0' no caso contrário.

10 – Processo de funcionamento do sensor de performance, de acordo com a reivindicação nº 6, caracterizado por não sinalizar quaisquer variações na entrada do detector de estabilidade (11) que ocorram durante o semi-ciclo negativo do sinal de relógio (5), para o caso de um Flip-flop com activação pelo flanco ascendente do relógio, ou que ocorram durante o semi-ciclo positivo do sinal de relógio para o caso de um Flip-flop com activação pelo flanco descendente do relógio.

11 – Processo de funcionamento do sensor de performance, de acordo com a reivindicação nº 6, caracterizado por utilizar uma banda de guarda virtual que permite identificar as transições tardias na entrada de dados do

Flip-flop (6), sendo esta banda de guarda criada pelo atraso introduzido pelo elemento de atraso (3), sendo este atraso dependente das variações do processo de fabricação, da tensão de alimentação, da temperatura e da degradação provocada pelo envelhecimento existente no circuito.

Faro, 17 de Abril de 2012

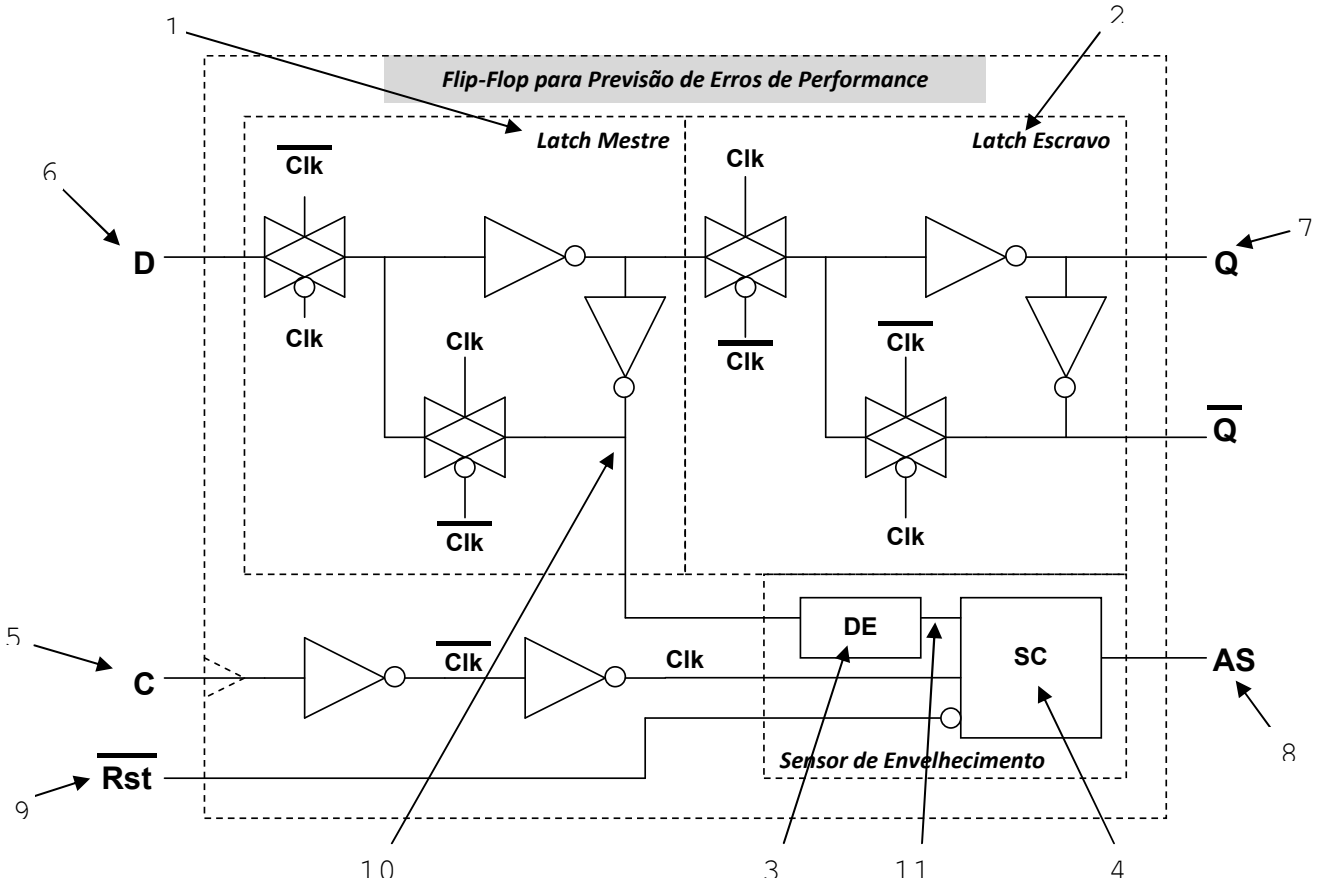
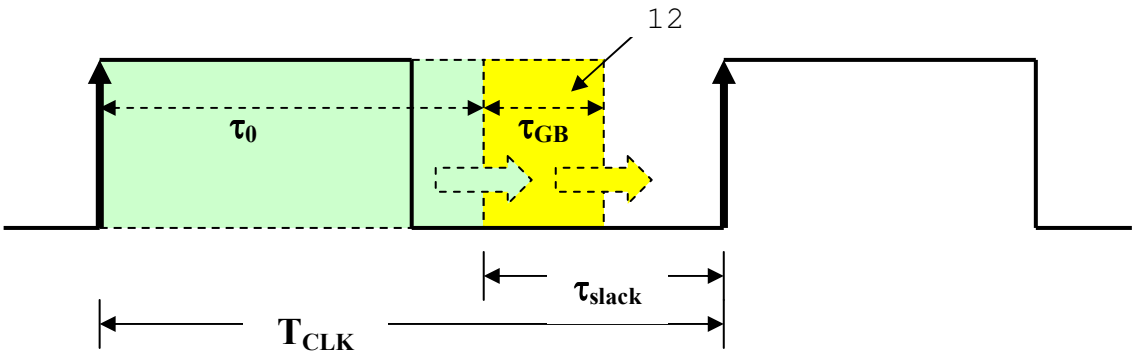
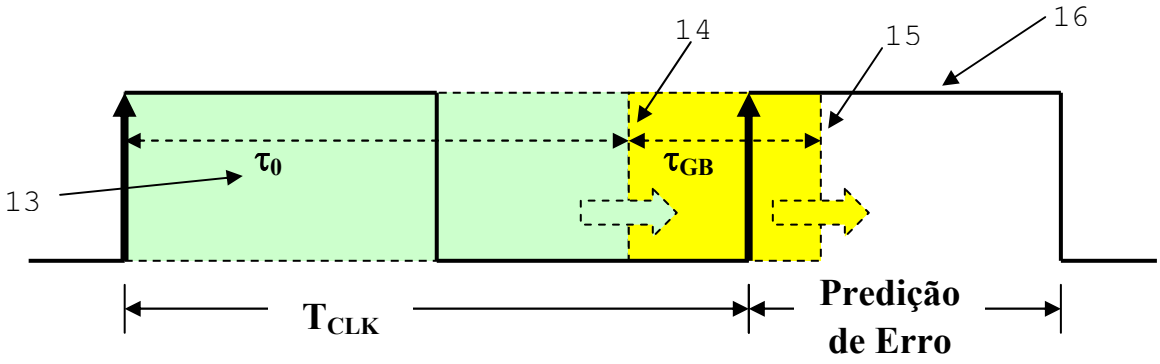

1
2
Flip-Flop para Previsão de Erros de Performance
Latch Mestre
Latch Escravo
6
D
Clk
Clk
Clk
Clk
Clk
Clk
Clk
Clk
Q
7
Q
5
C
Clk
Clk
DE
SC
AS
8
Sensor de Envelhecimento
9
Rst
10
3
11
4

Fig.1



(a)



(b)

Fig.2

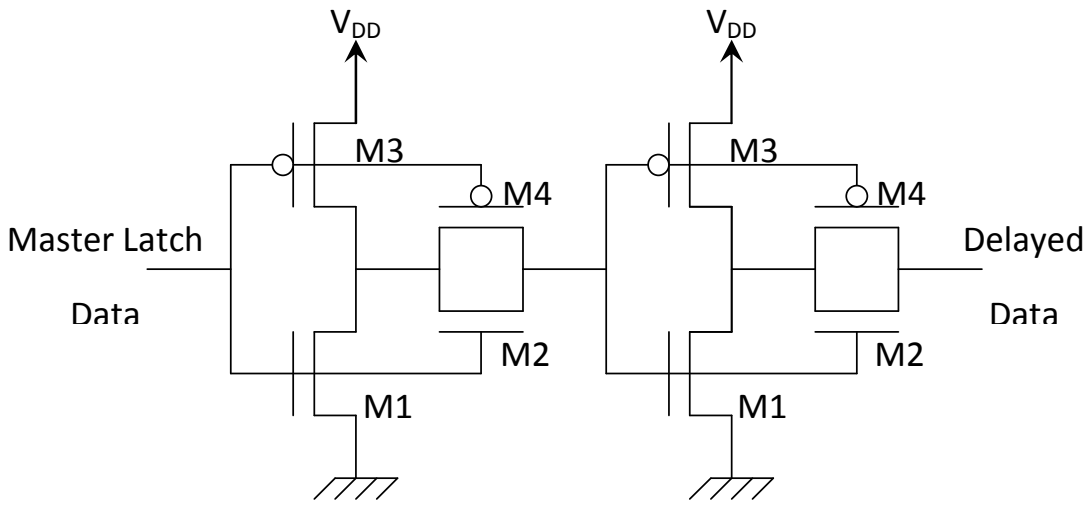

V_{DD}
V_{DD}
M3
M3
M4
M4
Master Latch
Data
Delayed
Data
M2
M2
M1
M1

Fig. 3

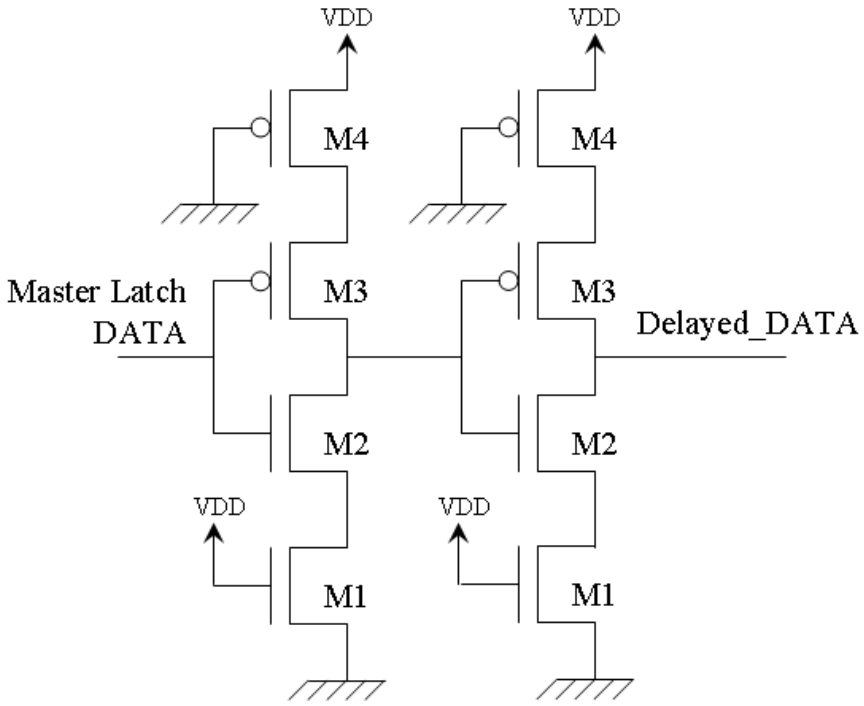

VDD
VDD
M4
M4
Master Latch
DATA
M3
M3
Delayed_DATA
M2
M2
VDD
VDD
M1
M1

Fig.4

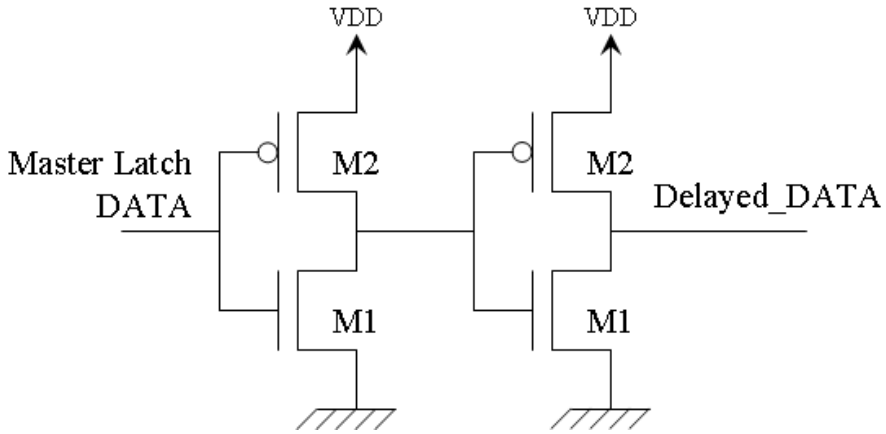

VDD
VDD
Master Latch
DATA
M2
M2
Delayed_DATA
M1
M1

Fig.5

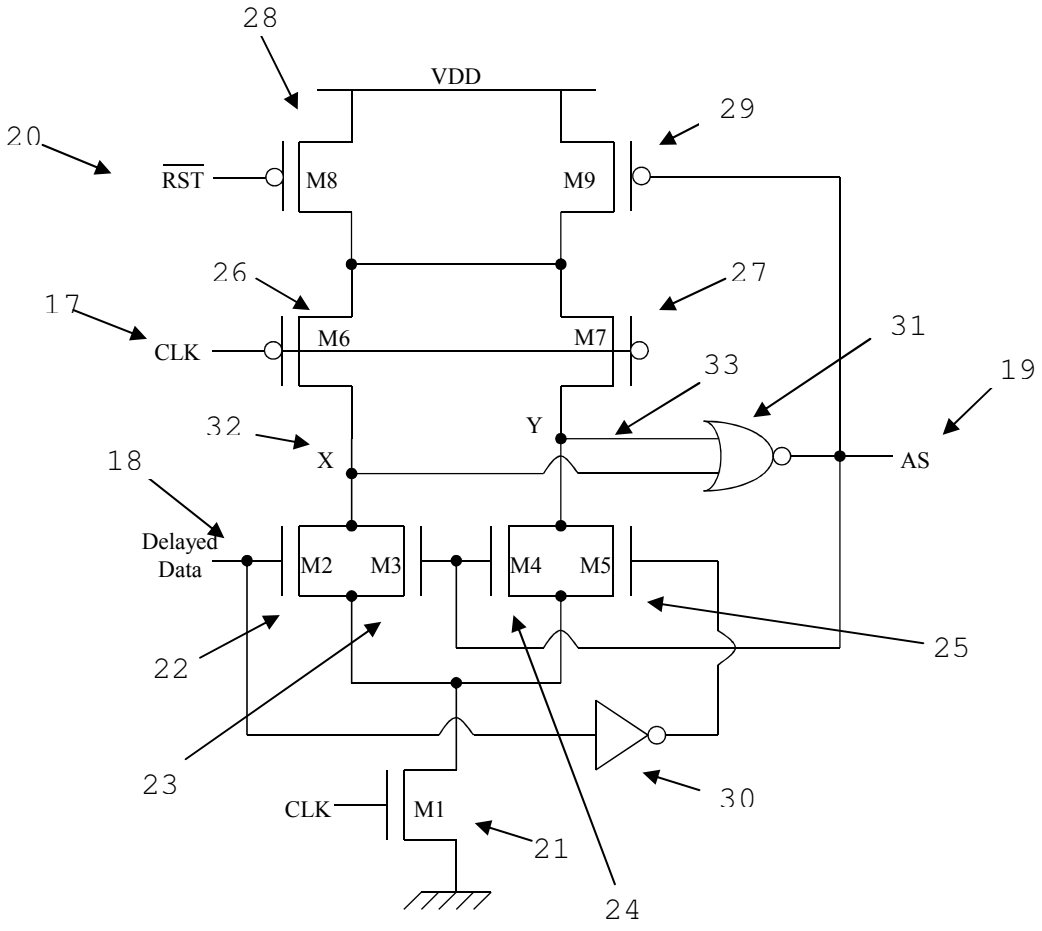

28
VDD
20
29
RST
M8
M9
26
27
17
CLK
M6
M7
31
33
19
32
Y
X
AS
18
Delayed
Data
M2
M3
M4
M5
22
25
23
CLK
M1
30
21
24

Fig.6