Technical University of Munich
TUM School of Computation, Information and Technology

Sensors for Data Bus Protection

Seyed Hamidreza Moghadas Tabatabaei Zavareh

Complete reprint of the dissertation approved by the TUM School of Computation, Information and Technology of the Technical University of Munich for the award of the Doktor der Ingenieurwissenschaften (Dr.-Ing.).

Chair:	Prof. Dr. Wolfgang Utschick
Examiners:	

1. Prof. Dr.-Ing. Georg Sigl
2. Prof. Dr. Ralf Brederlow

The dissertation was submitted to the Technical University of Munich on 18.10.2023 and accepted by the TUM School of Computation, Information and Technology on 20.03.2024.

Abstract

Microprobing, as an invasive physical attack technique, poses a formidable threat to the integrity and confidentiality of on-chip signals, encompassing vital components such as data buses and communication lines. This intrusive method enables unauthorized access to highly sensitive information, ranging from firmware and on-chip memory contents to cryptographic keys and other critical data transmitted through physical metal lines and cells. To safeguard against the perils of microprobing, the deployment of robust countermeasures becomes imperative.

This research is dedicated to the development of novel probing detector circuitry, engineered to effectively detect and respond to microprobing attacks while minimizing design overhead. The proposed detector circuitry exhibits remarkable improvements in sensitivity, detection capabilities, yield, and reliability compared to existing approaches. Notably, the concept offers comprehensive protection for both equal and non-equal length bus lines, surpassing the limitations of conventional techniques. The design and analysis of the detector circuitry are specifically tailored for an advanced sub-40 nm industrial process node, ensuring compatibility and effectiveness in cutting-edge technologies.

The efficacy of the proposed concept is thoroughly evaluated, encompassing variations in process, voltage, and temperature, as well as statistical process distribution. Furthermore, the design incorporates robust measures to withstand the challenges posed by coupling and jitter noises, thereby fortifying the circuitry against potential disruptions. Special attention is given to mitigating the vulnerabilities associated with locking attacks on ring oscillators, further enhancing the security measures implemented.

It is paramount to address the issue of aging effects, particularly when dealing with the utilization of very low capacitance probing heads, such as those encountered in Focused Ion Beam applications or 20 fF microprobes. Aging mechanisms introduce significant risks to on-chip CMOS circuitry, manifesting during the High Temperature Operating Life (HTOL) test phase and throughout the lifetime of the circuitry. These mechanisms induce variations in propagation delay, necessitating the development of an aging built-in self-test with minimal overhead. The objective is to effectively detect and, ideally, predict the aging status of the circuitry, thereby facilitating the identification of potential points of wear and tear.

In this context, novel methods are proposed to alleviate the overhead associated with aging built-in self-tests. The foundation of these methods lies in exploiting the frequency dependence exhibited by aging curves. To facilitate the storage and retrieval of frequency dependence curve data and time-to-digital converter calibration data, a lookup table is employed, conveniently stored in an on-chip memory. As an alternative to encryption, a computationally less intensive approach is proposed: the utilization of a physical unclonable function to protect the aforementioned data. However, it is essential to acknowledge that aging can introduce alterations to the sign bits of the physical unclonable function, necessitating a comprehensive analysis of measures to enhance its reliability against the deleterious effects of aging.

By diligently pursuing these research directions, we aim to establish advanced security measures capable of mitigating the risks posed by microprobing attacks, while concurrently addressing the challenges associated with aging effects. Through the amalgamation of cutting-edge technologies and innovative methodologies, we strive to secure the confidentiality, integrity, and longevity of on-chip systems in the face of evolving threats.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my esteemed advisor, Professor Dr.-Ing. Georg Sigl from *Technical University of Munich*, for his unwavering support and invaluable guidance throughout the arduous journey of this dissertation. I am deeply indebted to Professor Sigl for his invaluable support and guidance, without which this journey would have been devoid of purpose and direction. His unwavering commitment to my academic and personal growth has been instrumental in shaping my path and reaching a meaningful destination. I express my sincere gratitude to Professor Sigl for his unwavering support, which has been essential in the realization of my goals and the culmination of this journey. His profound expertise and extensive knowledge in the field of smart cards have been instrumental in shaping my understanding of the development and testing strategies that yield optimal outcomes while considering cost implications. Professor Sigl's industrial background has provided a unique opportunity to delve into more realistic scenarios, and I am profoundly grateful for the invaluable learning experiences that have resulted from his mentorship.

I would like to extend my heartfelt appreciation to Dr.-Ing. Michael Pehl from *Technical University of Munich* for his exceptional supervision, particularly in the probing detection domain. Dr. Pehl's profound understanding of circuit design, physical unclonable functions, and side channel topics has been instrumental in shaping the development of the probing detector concept. I am deeply grateful for the extensive hours Dr. Pehl dedicated to our discussions, both within and beyond regular working hours, as they have significantly contributed to the refinement and organization of my research endeavors.

I would also like to express my sincere thanks to my Mentor Dr. Ulrich Fiedler from *Infineon Technologies Neubiberg* for his comprehensive support in all aspects of this research, with regard to time management and resource allocation. Without his support this journey had most likely not been complete up to the destination. Dr. Fiedler's support, understanding, and above all, patience, have been invaluable throughout the entire process. Dr. Fiedler's remarkable leadership qualities, combining strictness and support, are a testament to his exceptional character, making him a rare and highly respected figure in the industry. Moreover, Dr. Fiedler's extensive network within the industry has been instrumental in facilitating consultations and resolving specific challenges as they arose. I am deeply grateful to Dr. Fiedler for providing financial support for publications and the procurement of essential software licenses crucial for the success of this research endeavor.

I am grateful to Professor Dr. Doris Schmitt-Landsiedel from *Technical University of Munich* for her unwavering belief in my potential and her steadfast support throughout this journey. Her guidance and encouragement have been invaluable, inspiring me to strive for excellence and pursue my research aspirations. Thank you for your unwavering support and belief in me.

I am grateful to Professor Dr. Ing Georg Fischer from *Friedrich Alexander University of Erlangen-Nuremberg* for his invaluable support and contributions regarding the aging topic and one of the early publications. His expertise and guidance have greatly influenced my understanding of aging phenomena and their implications. I would like to express my appreciation for his emphasis on engineering principles and resource optimization. His teachings have taught me to prioritize power, area, and financial efficiency in my work.

I would like to thank my employer, Infineon Technologies for financially and technically supporting me during the dissertation.

Additionally in the process of performing this research, there have been fruitful on-going discussions with Michael Weiner from *BMW Munich*, Christian Pacha, Florian Seilmeier, Human Boluki, Liang Wenfeng, Michael Murkovic, from *Infineon Technologies, Neubiberg.*

I would like to thank students such as Ghazal Safian who partially supported this work. I would like to thank all other individual colleagues whose names might not be included, and/or unintentionally forgotten here, however have kindly assisted with my inquiries along the way.

To my loving wife, Hamta: In this journey through academia, your unwavering support and sacrifices has been my guiding light. This dissertation is a testament to your belief in me.

Last but certainly not least, I extend my heartfelt gratitude to my beloved parents for their unwavering support and encouragement throughout my academic journey. Their constant presence and guidance have been pivotal in shaping the path I have chosen to pursue. I am truly fortunate to have such loving and supportive parents who have instilled in me the values of perseverance, determination, and resilience.

My parents have played a significant role in my life, serving as a constant source of inspiration and motivation. Their unconditional love and belief in my abilities have propelled me forward, enabling me to overcome numerous challenges encountered along the way. Their unwavering support has given me the confidence to pursue my academic goals and strive for excellence.

I am profoundly grateful for their sacrifices, both personal and financial, which have made it possible for me to embark on this educational journey. Their continuous encouragement and belief in my potential have been instrumental in my achievements thus far.

In conclusion, I extend my heartfelt thanks to all those who have contributed to the realization of this dissertation. Your unwavering support, guidance, and encouragement have been instrumental in shaping my academic journey, and I am profoundly grateful for the invaluable lessons and experiences that have resulted from our collaborations.

“Being well organized is better than doubling profits! You should always keep the distant future in mind – that’s what matters most.”
(Werner von Siemens to his brother Carl, July 17/18, 1868)

Contents

Chapter 1

Introduction

1.1 Motivation

Microprobes play a crucial role as measurement heads employed for extracting signals from CMOS circuitry. While they serve constructive purposes such as facilitating debugging, they also possess the potential for destructive applications, including cloning or reverse engineering of Application Specific Integrated Circuits (ASICs). Consequently, the detection of microprobing attacks holds significant importance in advancing the physical security of ASICs. This research is centered around the development of a novel microprobing detection mechanism that aims to enhance the protection of ASICs against unauthorized access and malicious activities.

With the continuous advancement of deep-sub-micron processes in recent years, the issue of aging has emerged as an additional variation mechanism that impacts the performance and reliability of integrated circuits. Aging phenomena introduce uncertainties and complexities in accurately modeling and predicting their effects. Consequently, there is a pressing need to devise innovative in-situ or prediction circuitry that can effectively forecast, detect, and mitigate the adverse impact of aging with minimal computational overhead. By addressing the challenges associated with aging, the reliability and lifespan of ASICs can be significantly improved.

In addition to the detection of microprobing attacks and aging mitigation, the concept of aging monitoring holds promise in safeguarding on-chip circuitry, including the probing detector and its associated data bus. The monitoring of aging effects can enable the identification of potential vulnerabilities and degradation points, thereby allowing for proactive maintenance and repair. To protect the critical data required for aging monitors, alternative approaches such as the utilization of physical unclonable functions (PUFs) can be explored. PUFs offer a computationally less intensive method for safeguarding memory contents compared to traditional encryption techniques. However, it is important to note that PUFs themselves are susceptible to aging, necessitating the development of strategies to enhance their reliability and resilience against aging effects.

This study aims to comprehensively investigate and propose strategies for efficient aging modeling, prediction, and detection in the context of microprobing attacks. Furthermore, the integration of PUFs will be explored as a means to enhance the security and integrity of critical data associated with aging monitoring. By addressing these challenges, the overall reliability, robustness, and physical security of ASICs can be significantly strengthened, providing a solid foundation for their widespread adoption in various applications.

1.2 Overview and Structure

This thesis aims to address several key aspects related to the physical security and aging effects in CMOS circuitry. The structure of the thesis is organized as follows:

Aging in CMOS Circuitry

In chapter 2, the primary mechanism underlying aging in CMOS circuitry is comprehensively discussed. The inherent limitations of existing aging sensors are then examined, highlighting the need for more robust and cost-effective sensor circuitry. Additionally, a novel concept for protecting on-chip circuitry, including the probing detector, is proposed.

Improving Resilience of Physical Unclonable Functions

In chapter 3, a method to enhance the resilience of a ring oscillator physical unclonable function against the adverse effects of aging is presented and rigorously analyzed. This physical unclonable function holds the potential to protect the memory of the aging sensor, thereby bolstering the overall security and reliability of the system.

Enhancing Security Measures Against Microprobing Attacks

Chapter 4 focuses on the core aspect of the research, namely the proposal and analysis of a novel probing detection concept specifically tailored for regular bus lines. The efficacy and performance of the proposed concept are thoroughly examined, considering various parameters and scenarios.

Probing Detection for Irregular Bus Lines and Calibration Mechanism

Building upon the foundation laid in chapter 4, chapter 5 delves into the probing detection concept for irregular bus lines. Furthermore, a calibration mechanism is introduced to enhance the Figures of Merit for the probing detection concept. Various second-order effects, such as jitter, are also discussed, and the influence of measurement time on the overall detection performance is explored. Comparative analyses with state-of-the-art probing detectors are conducted, culminating in a conclusive assessment. Additionally in this chapter a claim asserting the best predictability for equal length bus lines and the worst-case detection scenario involving a probe attached to the line with higher capacitance is substantiated. This provides further insight into the underlying principles and factors influencing the probing detection concept.

Summary and Conclusions

In chapter 6, a comprehensive summary of the work presented in this thesis is provided. The key findings, contributions, and implications of the research are summarized, offering a concise overview of the advancements made in the field of probing detection and aging mitigation in CMOS circuitry.

Copyright Notes

In Chapter 6 and in accordance with the guidelines set forth by the Institute of Electrical and Electronics Engineers (IEEE), this appendix serves to notify the inclusion of a copyright declaration within this thesis. The purpose of this declaration is to ensure proper acknowledgment and protection of intellectual property rights.

By adhering to the prescribed copyright declaration guidelines, this thesis upholds the principles of integrity, attribution, and respect for the contributions made by other researchers and authors in the respective fields of study.

This declaration stands as a testament to the commitment of the author and the academic institution to uphold ethical standards and promote responsible scholarly practices.

The specific details of the copyright declaration, including the relevant statements and notices, are documented in this appendix for transparency and compliance with the requirements of the IEEE.

By systematically addressing these various aspects, this research endeavors to contribute to the advancement of physical security and aging resilience in CMOS circuitry, fostering a more robust and reliable foundation for modern integrated circuit design and deployment.

Chapter 2

Aging in CMOS Circuitry

2.1 Introduction

Problem Statement

Chip production in deep-sub-micron technologies is facing different types of variations. Aging as one of the root causes for this variation, has been recently getting more attention and is regarded not only as a lifetime accumulative variation mechanism, but also as a dynamically short-term one [1]. Aging causes the deep-sub-micron technologies to change the propagation delay of digital and analog paths implemented by that process. The effect is dependent on the stress time, and stress frequency.

The growing significance of aging in application-specific integrated circuits (ASICs) underscores the need for the development of effective strategies to detect and differentiate aging effects from other forms of variation, while minimizing the impact on the die size. By effectively addressing the challenges associated with aging detection and differentiation while optimizing die size utilization, ASIC designers can enhance the longevity, reliability, and performance of their integrated circuits. This proactive approach enables the mitigation of aging-related issues and ensures the sustained functionality of ASICs throughout their operational lifetimes.

Aging occurs due to deep-sub-micron fabrication technology aging effects such as Negative Bias Temperature Instability (NBTI), Positive Bias Temperature Instability (PBTI), Hot Carrier Injection (HCI), and Time-Dependent Dielectric Breakdown (TDDB) [42] [49] [43] [12] [32] [16] [13] . Negative-biased temperature instability (NBTI) [48], [42], [25], [14], [21], [56], [17], [6], [57], [10], is one of the key aging reliability issues in CMOS technologies. It affects the PMOS performance by increasing its threshold voltage V_T over time, and consequently decreasing the drain current. The mechanism occurs due to localized traps in the gate insulator capturing charges from the channel [29]. On the circuit level, the aging issue degrades timing performance. As a result, critical paths may exceed the clock period causing timing failures.

It is worth it to carefully monitor and further handle aging due to its long [46] and short-term [1] effects on both digital and analog circuitry. One direct way to counter the deterioration of aging is to set more pessimistic timing margins in design time. This - the so-called pessimistic design - reduces the performance due to reduced clock frequency. As an alternative - or parallel to the margins - aging monitoring concepts are applied in the field [11].

The traditional Vernier time-to-digital converter-based BISTs present challenges in terms of calibration overhead, post-processing complexity, and aging stress on critical components. However, through a reduction in overhead based on the frequency dependence of the NBTI aging mechanism, significant improvements can be achieved in terms of area and power consumption. By embracing this approach, the semiconductor industry can enhance the reliability and performance of ASICs, ensuring their continued success in the face of aging-related challenges.

2.2 State of the Art

Replica Monitors

A replica monitor [15] is used as a method to figure out an average aging estimation of paths - on the same die that the replica is placed on - without having to directly measure those paths. These replica paths - specifically replica ring oscillators inserted on the same die - are used to monitor the aging effects which are based only on the on-time of the chip. This has the drawback that the result is dependent on the oscillating frequency of that oscillator, which could be under a pretty different stress trajectory compared to the single critical paths distributed all over the different circuits, and all over the die.

The aging level of the replica oscillators is suffering from an error compared to the main critical paths. This is firstly due to the non-matching bit patterns, in comparison to what the main paths get. Critical paths are stressed with different input bit patterns based on the in-field application. In contrast, a replica ring oscillator oscillates with a constant toggling rate determined out of its logic depth. Secondly - as another difference - a replica oscillator is made from a fixed number of gates. In contrast, each of the critical paths includes non-identical gate numbers and gate types.

Due to the mentioned differences, the helper data out of the replicas -emulating aging effects-, incorporate an error. This error is non-constant and varies from application to application, or from case to case.

A replica monitor is shown in Figure 2.1 [15]. A control unit takes care of measurement starts and stops, as well as the path selection process. A counter measures the frequency out of the ring oscillator, corresponding to the delay of the selected path. A post-processing unit reads the results out of the counter and translates them to correspondingly delayed information. It could be observed that the path inside the ring is selected to be out of either NANDs, NORs, adders, path gates, or a wire. Wires are used to analyze the effect on a metal line together with a single inverter ring oscillator. As a wire does not include an oxide layer, it is hence not endangered by NBTI. The inverter however in the feedback path is endangered. Other paths include more active elements to be measured. One reason for having these select paths is to measure the aging effect on the different types of logic, hence to gain more data regarding different types. The real application could be still very different from what is pre-designed, not only from a netlist perspective but also from an applied stress pattern perspective.

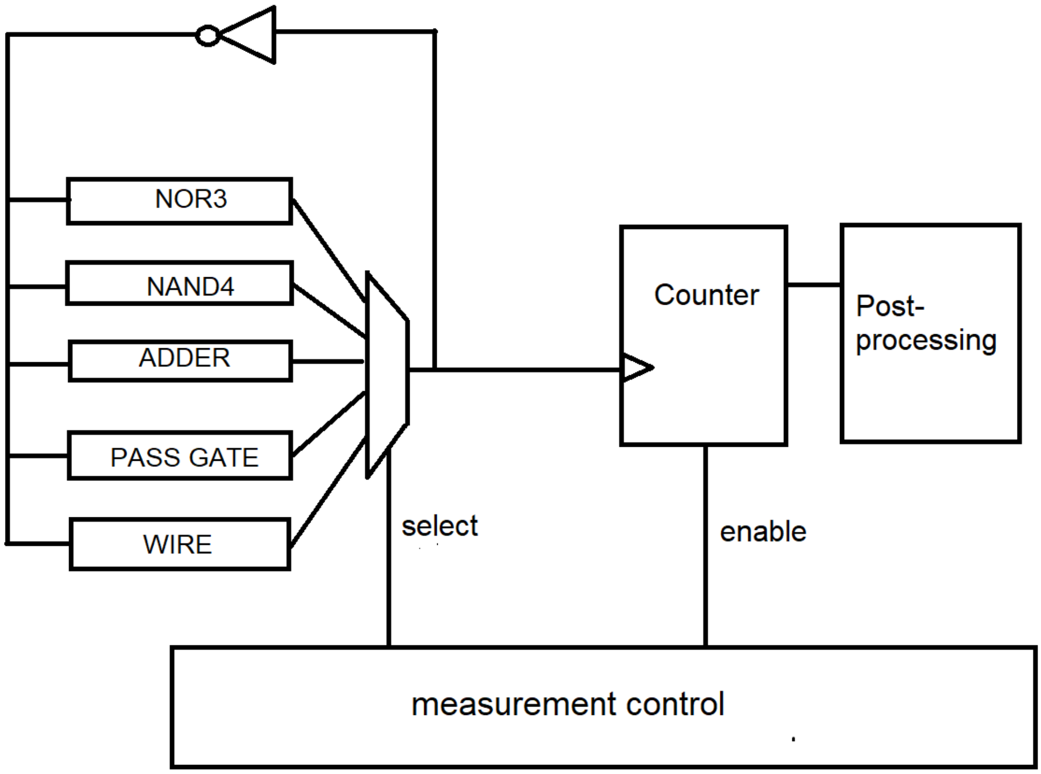


Figure 2.1: A simplified structure of a replica ring structure

In-situ monitors

In-situ monitors [7] [26] [59] [35] [55] [54] [60] measure the timing within the circuit. They are usually installed in the operation circuit, directly detecting timing errors.

The simplified structure of in-situ monitors - installed on several system critical paths - is shown in Figure 2.2. The core of such a system is a start-stop time to digital converter (TDC). The time to digital converter is used to measure the propagation delay inside the critical path.

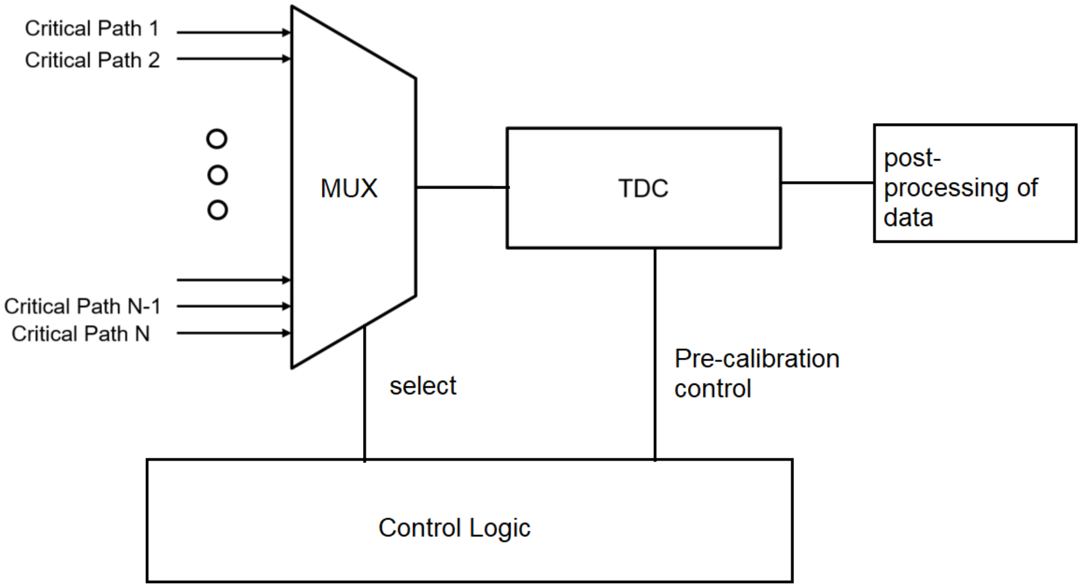


Figure 2.2: System-level implementation of aging in-situ monitors which share a central time to digital converter between several system paths to save space [14].

A time to digital converter measurement path is shown in Figure 2.3. There is a race condition between the start and stop rails, which is used to measure the propagation time. Please note that the $t1$ and $t2$ are different sizes. The difference in sizing is the basis for generating the timing steps.

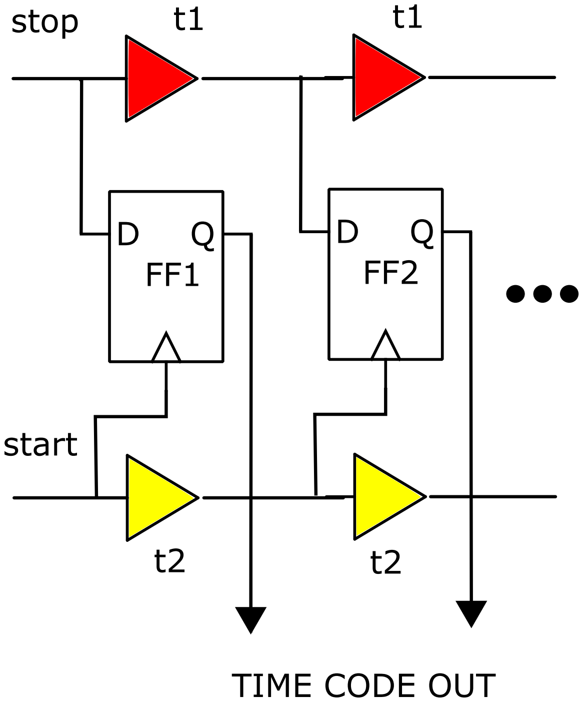


Figure 2.3: Structure of a vernier delay line TDC. The dual rails are subject to different aging trajectories.

2.3 Limitations with SoA and Observation "Least Toggling Bit Suffices"

Overhead, Routing Congestion, and MUX Size Problem

In-situ sensors measure the delay of a path, which uses a time to digital converter as a measurement core to monitor the delay of the aged Device Under Test (DUT). A time to digital converter requires calibration and post-processing blocks to protect it against the process, voltage, temperature, and noise variation.

A time to digital converter together with its calibration and post-processing circuitry is an area-consuming element. Hence it is expensive to build several units of it on a single SoC. Hence as a saving approach, one would like to use as little time to digital converter as possible in an Application Specific Integrated Circuit (ASIC).

Multiplexing the time to digital converters to a large number of different system paths all over the die, is an erroneous solution as big active multiplexers are required between the system paths and the measurement time to digital converter block. This results in a noticeable measurement error due to the huge MUX propagation activity delay. The bigger the MUX size gets, the higher the readout error, and hence the less reliable the measurement results become. The MUX size scales on the SoC top level by increasing the number of DUTs which are connected to the central time to the digital converter through the MUX.

Moreover, connecting the time to digital converter to several system paths through multiplexers requires a long test run time due to the lack of parallel operations within a multiplexed approach, which decreases the system performance. Meanwhile, connecting the time to digital converters to a large number of system paths requires additional metal routing to those paths, which increases the routing congestion in the physical design phase and must be prevented as much as possible.

The time to digital converter is itself an on-chip measurement device, whose measurement result is based on the delay of active on-chip cells, such as inverters and buffers. The same way that aging can affect an on-chip active component, so can it affect the on-chip measurement device as well. The time to digital converter includes dual rail on-chip delay lines, made out of inverter gates as the references. This is depicted in Figure 2.3. Each inverter gate is made out of one active PFET, and one active NFET. Hence the on-chip measurement devices are prone to an error in their measurement results which are age/stress-dependent.

The measurement results from the on-chip aging sensors, as well as any other time to digital converter-based sensors, are used to control and compensate the main variation mechanism. Such compensation could be performed by controlling a voltage, temperature, or frequency actuator. The problem occurs when the measurement result from the on-chip sensor is error-prone. Hence a wrong actuation process could take place. In case there is no actuation in place, the sensor results could be used to trigger a failure alarm.

The error-prone elements in a measurement time to digital converter are depicted in Figure 2.3. In a sub-gate delay time to digital converter, dual rail buffers sizes are not equal to each other, therefore each rail will have its own aging trajectory due to different oxide sizes. A time to digital converter sensor -which is the aging measurement core used at the system level approach- is itself not safe against aging due to the usage of active elements. In the case that the dual rails are from different size buffers, the dual rails are aged differently. This would negatively affect the reliability requirements of future SoCs. Reference [58] has analyzed and substantiated the claim that measurement errors in time to digital converter results occur due to aging. An aged time to digital converter is not a reliable measurement sensor for aging measurements, because it provides aging measurement results which reflect less aging than in reality. Hence an approach is required to reduce the stress on the time to digital converter, by reducing the number of required tests.

Aging sensors [14] are area/power consuming, when applied to many logical paths. Due to increased number of aging tests, these sensors are themselves prone to pop up false alarms due to aging inside themselves [58].

A 64-input-MUX time to digital converter based test chip with only one time to digital converter at [14] used 0.24 mm^2 of implementation area in a 45nm process. Hence it could be concluded that the real SoC top level case with thousands of Devices Under Test (DUT) distributed over much longer metal line distances - compared to that of [14] - and requiring reliability-certified Process, Voltage and Temperature calibrated time to digital converter- such as the ones in [9][18] - would require a huge amount of area/power for physical implementation. Up to 11% drift in a sensor result due to aging has been reported by [58], and hence needs to be addressed as a limitation.

Moreover, the state of the art is currently not providing prediction mechanisms and is based on a 100% detection approach by placing the sensor on the whole suspected data path. Imagine a sample case; for monitoring 100 Intellectual Properties (IPs), receiving data from a 32 bit bus -using a single shared central calibrated time to digital converter-, a huge MUX with 3200 inputs would be required. A delay of this big MUX is also measured by the central time to digital converter in addition to the main path delay, and is hence a big source for readout errors causing false alarms. The issue gets even bigger in the case that the MUX is non-symmetrical and is itself aging due to the high duty cycle.

An increase of the time to digital converter numbers to reduce the MUX size is not a suitable solution due to the overhead of the time to digital converter calibration and post processing area/power. An overhead limitation of the state of art, delay monitoring sensors in that they need a post-processing circuit. This post processing circuitry is explained in detail by [58][18]. The size of the post-processor has a direct correlation to the number of data paths which are under monitoring, as well as the time to digital converter code length, and PVT/noise calibrator order of sophistication. As the state of art at [14] is placing the sensors on every suspected stressed path, on a system level view, the post processing effort would be extremely high and the area overhead does not make economic sense. One might even decide to retain the older process rather than migrate, to avoid aging monitoring overhead in the case that the monitoring costs surpass the benefits of a migration process.

Observation: "Least Toggling Bit Suffices"

I substantiate a claim based on the results out of [56], [57] and also a highly cited work on NBTI in [10]. Results of these works, substantiate the claim of strong frequency dependency of NBTI aging.

Based on [56] frequency dependency of the NBTI-induced V_T shift for $> 100Hz$ has the following dependency with V_T at $f0$:

$$V_T(f) = V_T(f0) \cdot (\frac{f}{f0})^{-0.03323} \tag{2.1}$$

Where $f0$ is an initial frequency with e.g. $V_T(f0)$ threshold voltage, to be compared with f as any other frequencies in the range >100 Hz with threshold voltage of V_T.

The least toggling bit has the smallest frequency. Hence it suffers the most from V_T and delay shift, and is hence the most sensitive to NBTI aging.

2.4 Important special case MSB

Automotive Wheel Rotation Counter using a Micro-controller

This section provides a specific case example, to make the engineering concept more intuitive: Imagine that a micro controller working at a speed of 500 MHz contains a counter interrupt for counting the car wheel rotations as a feedback method coming from the car break system; the counter transfers the counted data onto a data bus to be delivered to the CPU. In this example the path under stress is from the counter input trigger pin to the capture register pin at the end of the data bus. A 32 bit counter most significant bit would toggle at

$$A = 2^{-31} \cdot 5 \cdot 10^{+8} = 0.23Hz \tag{2.2}$$

Based on [10], the NBTI effect for 110 °C is increasing by limiting to the zero frequency at <100 Hz range, but it remains almost constant at >100 Hz range, in comparison vs. the <100 Hz range. It could be concluded that this is the reason why less NBTI effect is reported in KHz, MHz or GHz ranges by other works such as [6]. Basically these ranges are where the least significant bits are mainly toggling. As the period is shorter in higher frequencies, the faster arriving recovery help, provided by the second half of each period of the higher frequency stress, can faster naturalize the V_T-increase caused by its first half period. In the 500 MHz 32 bit case, 10 most significant bits are located at <100 Hz range, carrying +95% of information entropy, and are hence harder affected by NBTI. A further increase in temperature would further increase the aging-oriented V_T-shift, because the temperature is inside the exponential term of NBTI function [50][28][20][21][27]. Adding a combination of other aging mechanisms on top of NBTI could further increase the PMOS threshold voltage.

A Communication IP for Multimedia Transfer

The application range covers the counters but is not limited to: Counters, analog to digital converters, digital to analog converters, digital communication modulator and demodulators based on FSK, ASK, PSK, OOK, QAM and GMSK schemes. The most significant bits are the bits with the smallest frequencies; e.g. of all of the mentioned modulation schemes, the analog waveform of the modulated digital signal is statistically with a higher probability continuous rather than discontinuous. Hence, with the digital representative of the modulated signal (at ADC output or DAC input), the least significant bits are statistically toggling with a higher rate compared to most significant bits. This can affect the SoC communication interface e.g. for the car radar or the IoT cloud communication.

Additionally, as the most significant bit contains 50% of modulation information entropy, it is the most sensitive bit to be protected against possible failures. For an ADC connected to a camera, statistically the gray scale variation has a higher probability compared to edge variation in the image [44], hence the least significant bits are toggling more than the most significant bits. A failure in the camera analog interface, could later affect the visual computing for an autonomously driven camera. The same happens to audio, temperature, and other analog inputs.

2.5 Concept

Please note that this approach is proposed for specific application cases where:

- The most-significant-bit of the bus, or a chosen replica most-significant-bit although not the main most-significant-bit, is toggling with the least frequency compared to the other bits.

In our approach, a redundant - and toggling with least rate most-significant-bit would be used to detect failure in the bus bits based on the NBTI frequency-dependency models. Any non-toggling most-significant-bit would be filtered out, to not affect the accuracy of the approach.

Generalization to "Least Toggling Bit"

We choose a data bus, with a toggling most-significant-bit - with the least frequency inside the bus bits - , as the base predictor for NBTI time to failure due to the following reasons:

1. The smallest toggling frequency, is the weakest point regarding NBTI and is hence used as a basis for prediction and detection. In several applications which are sending the data to the data bus, the most-significant-toggling-bit includes the smallest frequency.

2. It is stressed more regularly than other parts of the system on the top level SoC hierarchy, as the main communication infrastructure between the SoC top level

IPs. Top level data busses consist of several levels of buffer stages to improve the timing. The buffer delay value varies through the NBTI.

3. A top level data bus is one of the longest paths among all the data paths existing on the SoC, as it has to go through the whole chip top level. The data path involves an acceptable logic depth, because it is obligated to buffer the data for the whole SoC top level through several buffer stages. The propagation delay of the data path in bus, averages the delay variation over a larger area on die. The on-chip replica ring oscillator[28][20] - as an alternative aging measurement mechanism - is located only on one specific local corner on the SoC and hence can only measure that single corner. For measuring all the corners with replica ring oscillators, several of them are required to be placed all over the top level and averaged among using counter blocks, which is area and power consuming.

Identification of Highly Likely Toggling MSB

The choice of which most significant bit to use, is pretty much application dependent and can be decided on a specific case-by-case basis. In the wheel rotation counter case for a car, it could e.g. be figured out by averaging which most significant bit is toggling with the highest probability in a car driving at a normal speed. The same could happen e.g. in a frequency shift keying modulation scheme. The probability of the toggling rate is to be calculated out of emulator devices pre-tapeout, to generate a signal transition toggling pattern. "An emulator logs the switching activity in a switching activity interchange format (SAIF) file, also possible in a signal database file like FSDB or VCD [2]." The SAIF file in an application specific integrated circuits design flow, could therefore be the basis in finding out the most significant toggling bit.

In order to have a double check in field, one might want to add a checker IP to the bus. The task of such a checker IP is to check average toggling of the chosen MSBs, and to ensure that it is within the range predicted pre-tape out of the FSDB pattern.

Aging Alarm

An inverter stage as one of the bus line drivers, with delay d_i and load CAP_i is shown in Figure 2.4.

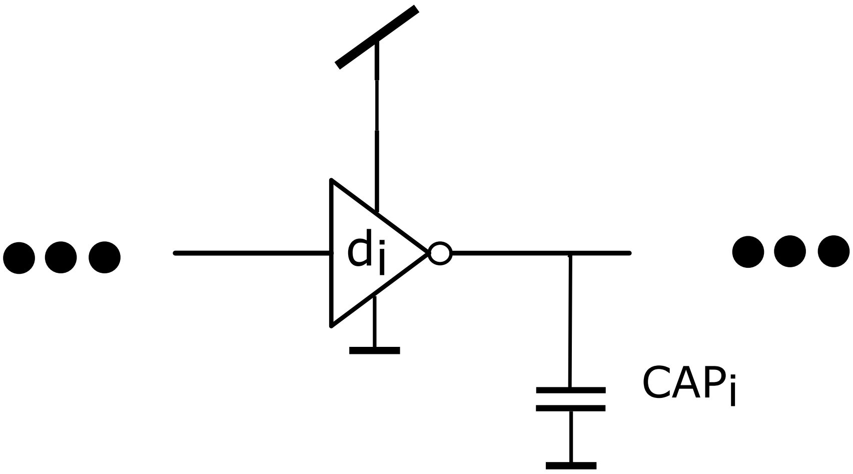


Figure 2.4: An inverter stage

The bus line delay is:

$$d = \sum_{i=1}^{M} d_i \tag{2.3}$$

with M the number of inverters in the line and d_i the delay per stage. The estimate for d_i based on the *alpha-power* model for transistors is [43] [12]:

$$d_i = k \cdot \frac{CAP_i \cdot V_{DD}}{(V_{DD} - V_T)^{\alpha}} \tag{2.4}$$

In this, V_{DD} is the supply voltage of the bus rails, CAP_i the output capacitive load of an inverter gate, V_T the threshold voltage, α the velocity saturation coefficient of the carriers, and k the trans-resistance [8].

A factor φ is defined by Equation (2.5), which is determined by the technology.

$$\varphi = k \cdot \frac{V_{DD}}{(V_{DD} - V_T)^{\alpha}} \tag{2.5}$$

Hence delay d_i is proportional to CAP_i, and is simplified as:

$$d_i = \varphi \cdot CAP_i \tag{2.6}$$

The line delay d gets increased due to process, voltage, temperature corners. Hence d_{varied} is calculated in sign-off phase together with the margins for process, temperature and voltage corners.

$$d_{varied} = d_{original} + \Delta d_{process} + \Delta d_{temperature} + \Delta d_{Voltage} \tag{2.7}$$

The least toggling bit has an additive reliability margin - added in sign-off phase-named as remaining slack: $RS_{sign-off}$.

$$RS_{sign-off} = T_{clk} - max(d_{varied}) \tag{2.8}$$

Where T_{clk} is period of the system clock, and $max(d_{varied})$ is the maximum delay, among all of the logic paths. Post sign-off phase the $RS_{sign-off}$ is optimized by the physical design, to be equal or bigger than zero.

The value of $RS_{sign-off}$ depends on physical design tightening conditions such as layout parasitic, applied clock frequency, corners, etc. $RS_{sign-off}$ is set at sign-off phase by static timing analyses, to cover the path reliability for the worst case design corner.

Hence post sign-off $RS_{sign-off}$ is not allowed to be violated on any of the design corners in the field. Hence an in-field calibration of the $RS_{sign-off}$ is not further required as it covers worst case regarding different scenarios.

In field, a timing sensor measures the actual slack of the least toggling bit, by comparing the delay of the logic path with that of the clock period.

$$RS_{field} = T_{clk} - d_{field} \tag{2.9}$$

where RS_{field} is the remaining slack in field, T_{clk} is clock period, and d_{field} is actual delay of path after aging in the field.

Due to NBTI aging, threshold voltage V_T is varied based on the cumulative effect of stress frequency applied to the MOSFET over its lifetime. V_T^{aged} is the threshold voltage after aging, V_T^0 is the threshold voltage before aging, and f is defined as toggling frequency of the signal.

$$V_T^{aged} = V_T^0 + \Delta V_T(f) \tag{2.10}$$

hence d_{field} is also varied based on:

$$d_{field} = \sum_{i=1}^{M} d_i = \sum_{i=1}^{M} k \cdot \frac{CAP_i \cdot V_{DD}}{(V_{DD} - V_T - \Delta V_T(f))^{\alpha}} \tag{2.11}$$

hence :

$$RS_{field} = T_{clk} - \sum_{i=1}^{M} k \cdot \frac{CAP_i \cdot V_{DD}}{(V_{DD} - V_T - \Delta V_T(f))^{\alpha}} \tag{2.12}$$

An alarm is triggered when the actual slack in the field gets lower than per sign-off-designed $RS_{sign-off}$. As $RS_{sign-off}$ has been determined for the worst case corner at sign-off phase, it is valid for all design corners, and hence is valid without further calibration.

$$ifRS_{field} < RS_{sign-off} \Rightarrow alarm = on \tag{2.13}$$

2.6 Evaluation of Benefits and Limitations

Benefits

The proposed approach of least-toggling-bit-based (or, under valid conditions, most-significant-bit-based) reduction of aging tests offers several significant benefits:

1. **Reduction of TDC Measurement Error:** By reducing the stress on the central TDC sensor, the proposed approach effectively minimizes the aging-oriented measurement error associated with the TDC.

2. **Lower Number of TDC Sensors:** The significant reduction in the number of required tests translates into a considerably lower number of needed TDC sensors, thereby reducing the overall complexity and resource requirements.

3. **Decreased MUX Overhead:** The approach leads to a reduction in the size of the MUX, resulting in decreased overhead and mitigating the measurement error stemming from MUX size and MUX aging effects.

4. **Reduced Area and Routing Congestion:** By reducing the metal routing, the proposed approach effectively decreases the overall area and routing congestion, resulting in improved layout efficiency and reduced design time.

5. **Total Area and Power Reduction:** The combined effect of the aforementioned benefits contributes to a significant reduction in the total area and power consumption of the system.

In contrast to current state-of-the-art approaches that advocate measuring each potentially critical path, the proposed approach focuses on utilizing only the least toggling bit within the data bus. This approach leads to a substantial reduction in the number of required measurements, as well as a decrease in the required number of time-to-digital sensors, MUX size, and metal routing. Since the selected bit is already present within the SoC's data bus, it does not introduce additional overhead. Consequently, this approach eliminates the need for implementing multiple replica ring oscillator structures throughout the SoC as replica monitors, resulting in significant area savings compared to the state-of-the-art methods.

Furthermore, this approach offers notable power savings, which is particularly crucial for portable battery-powered devices and Near Field Communication (NFC)-based consumers like NFC smart cards. It also addresses power concerns in non-portable IoT SoCs due to the effort required to minimize IR-Drop. Therefore, the proposed approach significantly reduces power consumption compared to existing aging sensors.

The overhead saving achieved by this approach can be estimated using the bus width W. Since one bit out of W bits is measured, the percentage of overhead saving can be calculated as:

$$\text{Saving} = \left(\frac{W-1}{W}\right) \times 100\% \tag{2.14}$$

For example, with a 128-bit bus, the saving is equal to $\frac{127}{128} \times 100\% = 99.2\%$.

Limitations

Despite its numerous benefits, the proposed approach has certain limitations that need to be considered:

One limitation is related to identifying the correct least toggling bit among the most significant bits to use. In cases where the most significant bit does not toggle at all, it cannot serve as the basis for this approach. In such scenarios, the first toggling most significant bit with the least frequency is selected as the basis. This limitation is addressed by considering specific cases where the most significant bit exhibits the lowest toggling frequency. While a toggling most significant bit is required, it may not necessarily be the first one in the bus.

Chapter 3

Analyses and Reduction of Variation of Ring Oscillator PUF

3.1 Introduction

Reliability and robustness are paramount in the realm of Internet of Things (IoT)-cloud-based communication, playing a crucial role in the prospective development of the IoT concept [31]. Ensuring a secure and dependable identification of IoT devices is a fundamental requirement in this context. To address this need, the Ring Oscillator Physical Unclonable Function (RO-PUF) has emerged as a promising alternative solution for key storage, offering distinct advantages over traditional encrypted memories [45][1].

As an alternative to traditional encryption techniques, a computationally less intensive approach is proposed, namely the utilization of a Physical Unclonable Function (PUF) as a means to protect sensitive data. However, it is crucial to recognize that aging phenomena, as well as voltage and temperature variation can introduce significant alterations to the sign bits of the PUF, thereby warranting a comprehensive analysis of measures aimed at bolstering its reliability and resilience against the deleterious effects of aging.

In light of this consideration, it becomes imperative to develop robust strategies and mechanisms to mitigate the impact of aging as well as voltage and temperat-

[1] This chapter partially follows the author's previous publications [37, 38].

ure variation on PUFs. By conducting an in-depth examination of the underlying variation mechanisms and their potential consequences on the stability and performance of PUFs, we can identify effective countermeasures to enhance their long-term reliability. This analysis encompasses a range of aspects, including but not limited to characterization, modeling, and compensation techniques tailored to address the unique challenges posed by the variation.

The goal of this research is to advance our understanding of the variation effects on PUFs and devise innovative solutions to ensure their continued operation and effectiveness in real-world scenarios. By adopting a comprehensive approach and leveraging insights from various disciplines, we aim to design PUF-based systems that exhibit heightened resistance to the uncertainties introduced by this variation. Through these efforts, we can establish PUFs as a viable and dependable alternative for safeguarding sensitive data in a wide array of applications, from secure communication protocols to authentication mechanisms.

In order to gain a comprehensive understanding of the variation present in Ring Oscillator Physical Unclonable Functions (ROPUFs), it is crucial to acknowledge that the embedded ring oscillators themselves are not immune to the influence of various sources of variation. These sources encompass a wide range of factors, including temperature fluctuations, power supply instabilities, and the effects of aging in deep sub-micron regimes [50, 28].

Neglecting to adequately compensate for these factors can have detrimental consequences on the performance and reliability of ROPUFs. One particular manifestation of such failure is the occurrence of frequency drifts in the ROPUF oscillators, which serves as the primary focus of this chapter. By examining and addressing the underlying causes of frequency drifts, we aim to enhance the robustness and stability of ROPUFs in the face of diverse environmental and aging-related variations.

To shed light on the intricacies of this variation and its impact on ROPUFs, a comprehensive analysis is undertaken. This analysis delves into the intricate interplay between the aforementioned sources of variation and their effects on the oscillation behavior of ROPUFs. By elucidating the underlying mechanisms and identifying the key factors contributing to frequency drifts, we can devise effective compensation techniques and mitigation strategies to ensure the reliable and accurate operation of ROPUFs in real-world applications.

Through meticulous characterization, modeling, and experimentation, we aim to unravel the complex relationship between variation sources and the performance of ROPUF oscillators. By gaining insights into the nature of frequency drifts and developing suitable compensation mechanisms, we can pave the way for the design and implementation of robust ROPUF-based systems that exhibit enhanced resistance to the detrimental effects of environmental and aging-induced variations.

In this chapter, I propose novel approaches to improve the hardware-based robustness of CMOS circuits against the deleterious effects of aging, and other CMOS variation effects. Additionally, a comprehensive spectral analysis is presented, specifically tailored for an implementation utilizing 65nm technology. By exploring these avenues, we aim to enhance the long-term reliability and resilience of IoT devices, thereby fortifying the overall security and performance of IoT-cloud-based communication systems.

3.2 Variation and Aging in Ring Oscillator PUF

ROPUF structure

In fig. 3.1, a ring oscillator Physical Unclonable Function is shown.

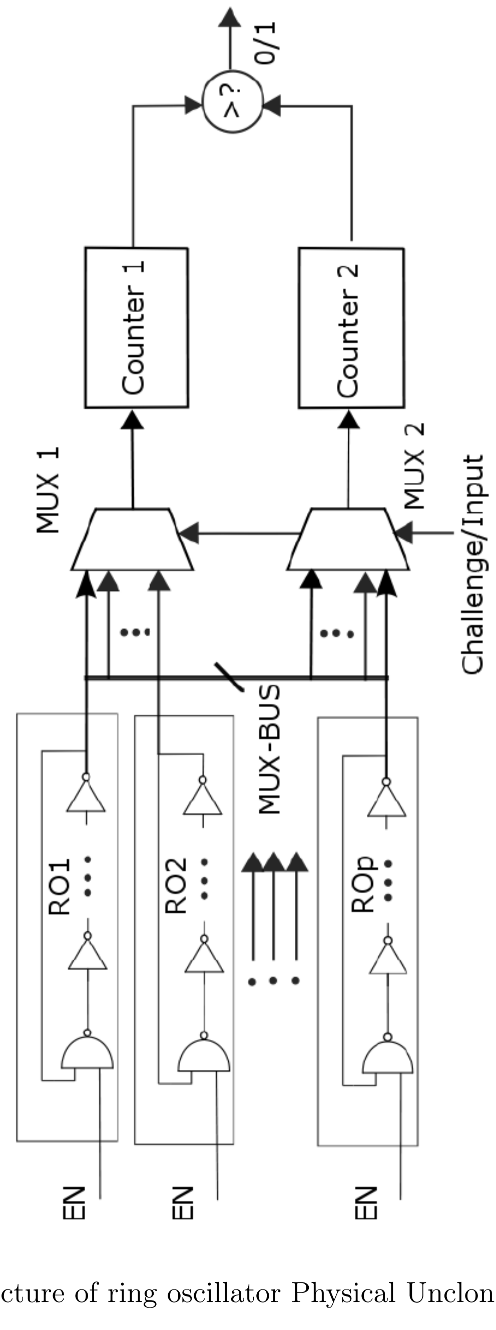


Figure 3.1: structure of ring oscillator Physical Unclonable Function

Each of the ring oscillator units oscillates at its unique frequency depending on the process variation of the specific area which the ring oscillator is implemented on. The control unit sends a MUX select word to the ROPUF MUXs which results in the selection of two of the ring oscillator frequencies out of the ring oscillator-set. Oscillations are counted by the counters and will be compared by the comparator to produce a random bit of 0 or 1. This process is named as one challenge. A binary random code word could be produced by applying continuous challenges to a conventional ROPUF described in [45].

On Necessity for Challenges with Forbidden RO Reuse

A challenge/response pair means that the two oscillators in a pair are compared using the two counters and a sign comparator. Based on which one has a bigger frequency, a response bit of either 0 or 1 is produced at the comparator output. This could be e.g. bit 1 if the counter 1 has a bigger value than counter 2. By sending several consecutive challenges, a stream of bits could be produced. For obtaining an unpredictable bit stream, each of the bits in the stream must remain completely unpredictable by a third party e.g. an attacker. According to [36], [53], it is forbidden to reuse one RO within more than one challenge. Imagine A, B,C and D are four oscillators. This means that a specifically paired set of oscillators A and B are always to be compared to each other. Accordingly, reusing the already paired oscillators to make new pairs - and hence building new challenges - is not allowed. This means that once A is paired with B, a new pairing of e.g. A to C or B to D is not further allowed. If this is not taken into account, an attacker could reconstruct the challenge involving A and B, without running the challenge. An attacker could run the challenge involving A and C once, and run the challenge involving B and D once. Comparing the results from the later two, the outcome of the challenge involving A and B is revealed. Imagine that three challenges would be used to make a single or multiple bit streams. Having the response bit of challenges A/C and B/D, the response of A/C would already be known without even submitting the challenge. Hence the response bit out of the A/C challenge is no longer unpredictable.

Aging

[30] substantiated the claim that aging increased bit flips in ROPUF, and proposed aging-robust ROPUF by reducing degradation speed of case-specific CMOS aging mechanisms in the ring oscillator MOSFETs, however does not consider fundamental bandwidth tuning of ROPUF. [33], [41] analyzed accelerated aging effect on Field Programmable Gate Array (FPGA)-ROPUF, however do not analyze and improve CMOS-ASIC based bandwidth design of ROPUF.

Aging and other types of variations in deep sub-micron CMOS technologies causes the ring oscillator to include sidebands around the center frequency [24], [33]. This problem is depicted in fig. 3.2, fig. 3.3. Variations out of temperature and supply could drift the inverter delay to increase or decrease compared to its nominal value. Hence they drift the oscillator frequency, either to a higher or to a lower value compared to its nominal center frequency. Aging increases the threshold voltage of the MOSFEETs. Hence aging results in an ever increased delay observed over a lifetime in the inverter gates. Hence aging results in reducing the center frequency of oscillators over their lifetime stress.

Any possible sideband interference could cause a bit flip at the comparator output [30],[33]. This causes an inconstant unreliable sign bit. Therefore, an optimized ring oscillator Physical Unclonable Function spectrum is needed. This - so called "tuned" ROPUF - is subject of discussion in upcoming sections.

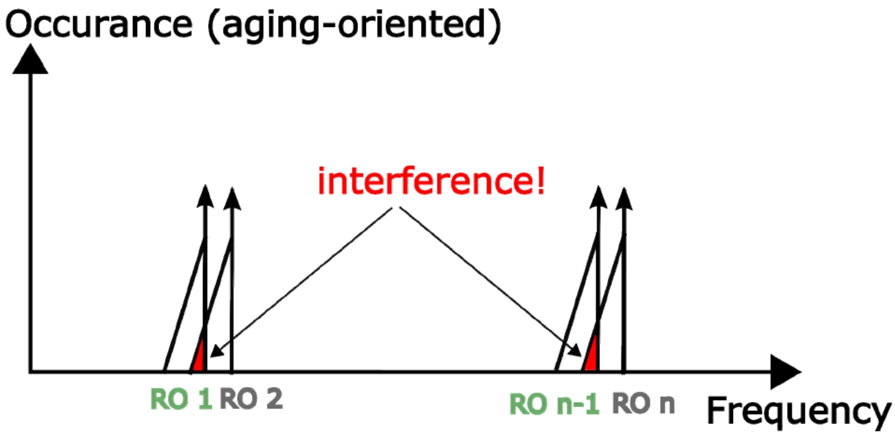


Figure 3.2: Aging oriented Bandwidth Interference

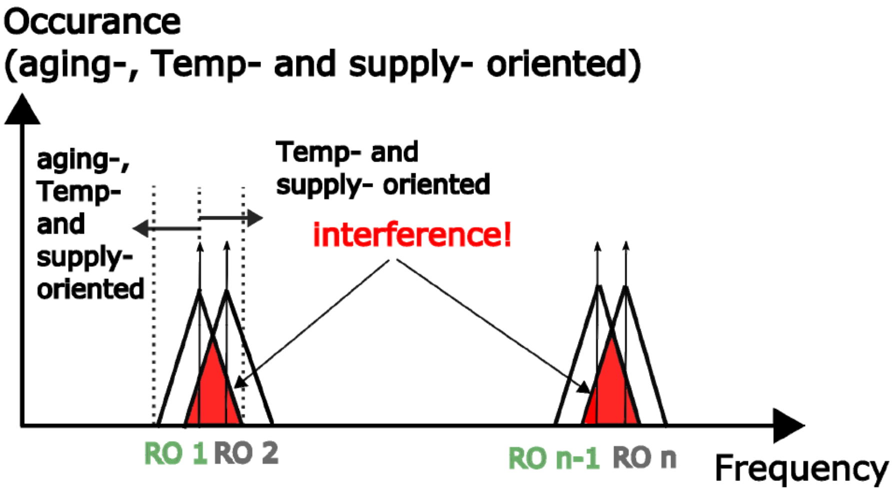

Occurance
(aging-, Temp- and supply- oriented)
aging-, Temp- and supply- oriented
Temp- and supply- oriented
interference!
RO 1
RO 2
RO n-1
RO n
Frequency

Figure 3.3: PVT-Aging oriented Bandwidth Interference

3.3 Tuning of Ring Oscillators: A Theoretical Model

Suppose an inverter stage in a ring structure, with the delay d_i and the load capacitance CAP_i.

The RO frequency is:

$$f = \frac{1}{2\sum_{i=1}^{M} d_i} \tag{3.1}$$

with M the number of inverters in the line and d_i the delay per stage. The estimate for d_i based on the *alpha-power* model for transistors is [43] [12]:

$$d_i = k \cdot \frac{CAP_i \cdot V_{DD}}{(V_{DD} - V_T)^{\alpha}} \tag{3.2}$$

In this, V_{DD} is the supply voltage of the bus rails, CAP_i the output capacitive load of an inverter gate, V_T the threshold voltage, α the velocity saturation coefficient of the carriers, and k the trans-resistance [8].

Changing the current of the current source, results in changing the V_{DD}, because V_{DD} is derived by

$$V_{DD} = R_{INV} \cdot I_{INV} \tag{3.3}$$

where R_{INV} is electrical resistance of the inverter gate, observed from its VDD supply pin. I_{INV} is the current flowing through the VDD supply pin.

$$\varphi = k \cdot \frac{R_{INV} \cdot I_{INV}}{(R_{INV} \cdot I_{INV} - V_T)^{\alpha}} \tag{3.4}$$

the RO's frequency f is

$$f = \frac{1}{2\varphi \cdot \sum_{i=1}^{M} CAP_i} \tag{3.5}$$

Changing the inverter load CAP_i results in adjusting the frequency. Changing the current source current, also results in changing the φ and hence changing the frequency. Both CAP_i and current source could be used as variables based on coefficients to tune the frequency.

A schematic of tuning for one of the inverter stages d_i is shown in fig. 3.4. An adjustable current source is used to supply the inverters. A variable capacitance network is used as a load for the inverter block. A set of coefficients are applied to the current source and switched capacitor bank to adjust the inverter delay. As the claim has been substantiated, adjusting the inverter delay adjusts the oscillator frequency.

$CAP_{coefficient}$ is the capacitance network coefficient used as a factor multiplied by a base capacitor value of $1fF$. Base capacitor value is named as CAP_{base}.

$$CAP_i = CAP_{coefficient} \cdot CAP_{base} \tag{3.6}$$

$I_{coefficient}$ is the current source coefficient used as a factor multiplied by a base current value of $200nA$. Base current value is named as I_{base}. The current factor multiplication occurs through a current mirror. Hence the $I_{coefficient}$ is applied as control values to the gate of the mirrors.

$$I_{INV} = I_{coefficient} \cdot I_{base} \tag{3.7}$$

The one stage delay is hence derived by:

$$d_i = k \cdot \frac{CAP_{coefficient} \cdot CAP_{base} \cdot R_{INV} \cdot I_{coefficient} \cdot I_{base}}{(R_{INV} \cdot I_{coefficient} \cdot I_{base} - V_T)^{\alpha}} \tag{3.8}$$

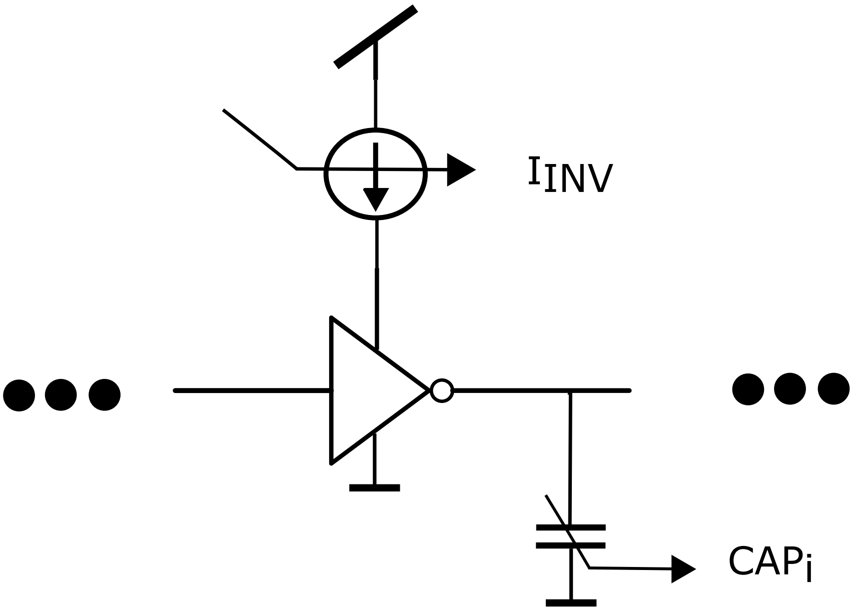


Figure 3.4: Tuning coefficient generation

Figure 3.5 shows implementation of the I_{INV} coefficients. I_{base} is mirrored through a current mirror with mirror factors of order 2^n. Through the coefficient bus, control of the current mirror is possible. Hence a binary formatted mirror current source is formed to generate the I_{INV}. The I_{base} could be implemented e.g. using a band-gap current reference [3].

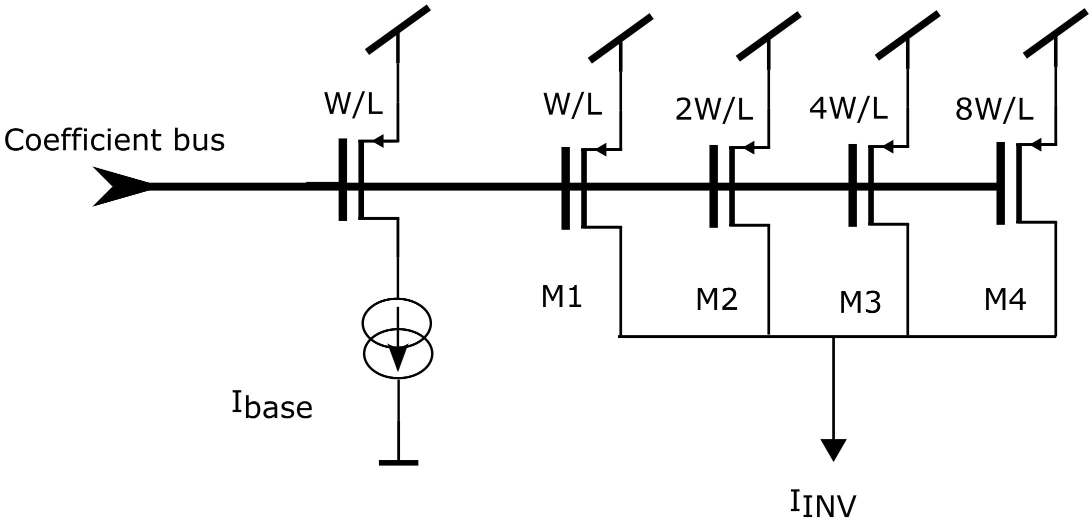


Figure 3.5: I_{INV} coefficients

Figure 3.6 shows implementation of the CAP_i. Factors of CAP_{base} are selected through the coefficient bus to implement CAP_i.

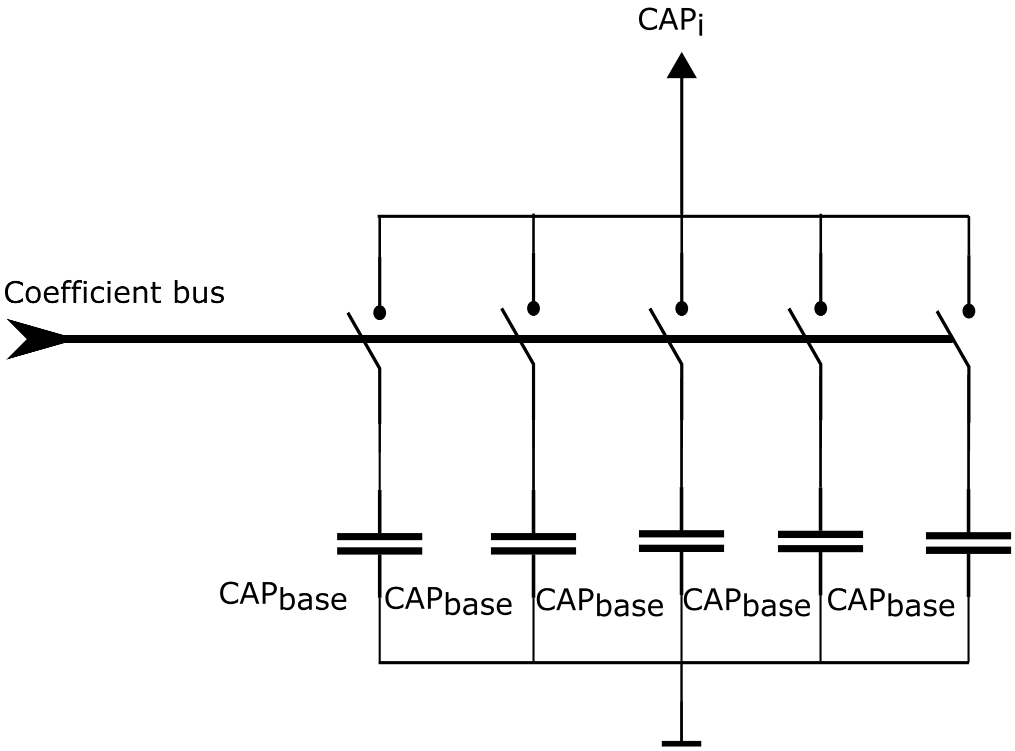


Figure 3.6: CAP_i coefficients

It will be further discussed in Section 3.4, how the coefficients are obtained for the coefficient bus.

On the Overhead

In reality it is enough to tune the oscillators based on a Drain Line Tuning (DLT) tune code, which is applied to the current source supply of the ring structure. Therefore to reduce the overhead, it is sufficient to only take the current-source-based tuning into account. A current source solution for supplying the ring structure, is usually used in practice - instead of the VDD-based source - for reducing the effect of VDD corner variation, and is thus not regarded as additional overhead. To reduce the number of the current sources, we could share two of the current sources between the ring structure pairs, and apply the pair specific DLT tuning code for each pair.

This has a benefit that the dynamic power usage of the ring structure is divided by a factor of "pair population = N/2", where N is the number of oscillators. Dynamic power usage of a ring structure is a practical issue in ring structure design, as the ring structure is switching in a very high frequency. Moreover, a parallel run requires not only a big enough power supply/battery, but also a big enough regulator output stage with enough stabilization to power all the parallel stages at the same time. Additionally it requires wide enough metal line routing on the die, to be able to overcome the IR-Drop and electromigration. Therefore running the oscillators in parallel is a very costly solution in terms of power and area. Hence running the ring structure in sequence - instead of in parallel - is in most of the cases "a must" in practice, specifically taking a battery powered application into account. This has however the drawback that not all the pairs could run in parallel, and higher run time would be required as the pairs have to run one after another one in a sequence. The sequential variation run time is multiplied by a factor of "pair population = N/2", compared to the parallel variation run time. Supposing that each oscillator needs 100 nanoseconds for stabilization and measurement, the overall run time would be equal to 50 microseconds e.g. for 500 pairs. This is a relatively short amount of time for generating the key and is regarded as an acceptable run time for a wide variety of applications. Based on the provided argument, two current mirrors are enough for supplying the ROPUF in a sequential form, with reduced power usage by the "N/2" factor, and with an acceptable run time.

A Discussion on the Frequency Domain Shifts

It is desired to reduce the overlap between the expectation of frequencies of two oscillators in a compared pair. This could be implemented by shifting the expectation of frequencies of the paired oscillators in a way, that the expectation of the frequencies, i.e. the side-bands, no longer suffer from interference. Hence the occurrence of an overlap will become less likely. Basically such a shifting should increase the distance between the expectation of frequencies of oscillators within a compared pair.

Suppose the frequencies from oscillators 1 and 2 - named as $f1$ and $f2$- are compared. For applying the frequency shifting, depending on the shifting direction, three different states could occur, given that aging is in place.

Increasing Both Frequencies

Suppose that, the tuning coefficients are given, so that both compared f1 and f2 increase, hence shifting both of the frequencies to the right side by increasing them. We would like to examine if this could improve reliability against aging, or if it is not effective. As could be seen in fig. 3.7 the overall shift is a superposition of the tuning effect - implemented by $CAP_{coefficient}$ and $I_{coefficient}$ - and the aging effect. The aging reduced the frequencies, however overall, both f1 and f2 are shifted to the right after taking aging into account.

It is observed that this tuning does not increase reliability, in the sense of reducing the overlapping of the ring oscillator spectrum, because after shifting to the right, the overlapping distance remains the same.

This tuning - being saved as coefficients on a non-secure, non volatile memory - does not provide a backdoor to an attacker, because both frequencies are shifted to the same -right- side. In the case that an attacker clones the coefficients, no info would be able to be obtained on which frequency is the higher one, hence the PUF output bit would not be obtained.

As this tuning does not gain or increase in reliability, it will not be further taken into account.

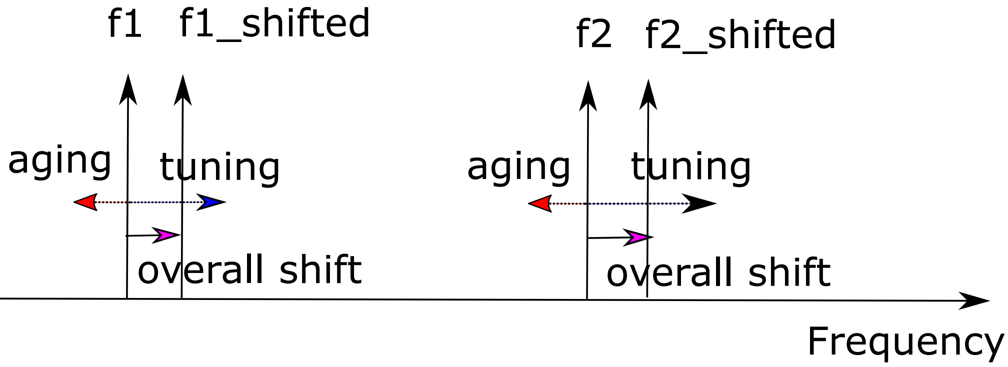


Figure 3.7: Tuning by shifting right both frequencies

Decreasing Both Frequencies

In this variant, the tuning coefficients are given, so that both compared f1 and f2 decreases, hence shifting both to the left side. As could be seen in fig. 3.8 the overall shift is a superposition of the tuning effect and the aging effect. Overall both f1 and f2 are shifted to the left after taking aging into account.

This tuning does not increase reliability, because after decreasing both frequencies, the distance between the expectation of frequencies remain in the same range as beforehand.

This tuning - being saved as coefficients on a non-secure non volatile memory - does not provide a backdoor to an attacker, because both frequencies are shifted to the same -left- side. In the case that an attacker were to clone the coefficients, no info would be able to be obtained on which frequency was the higher one, hence the PUF output bit would not be obtained.

However as this tuning does not gain or increase in reliability, it will not be further taken into account.

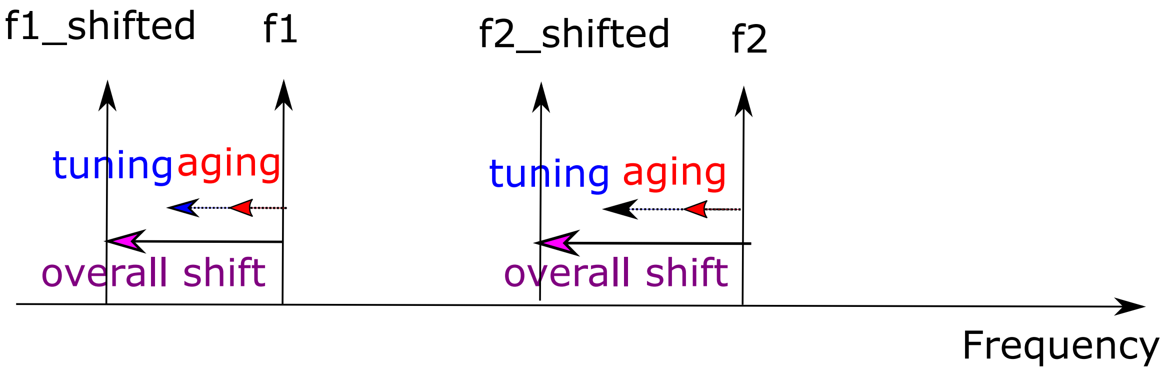


Figure 3.8: Tuning by shifting left both frequencies

Tuning Frequencies to Increase the Distance of Frequency Expectations

In this variant, the tuning coefficients are given, so that f1 decreases and f2 increases, hence shifting f1 to left and f2 to right. As could be seen in fig. 3.9 the overall shift is a superposition of the tuning effect, and the aging effect. Overall f1 is shifted to left, and f2 is shifted to right, after taking aging into account.

This tuning increases reliability, because it increases the distance between the expectation of f1 and f2, hence reducing the probability of interference between the frequency side-bands, hence reducing the probability of a bit flip out of their comparison. After the overall shift, the distance between the occurrence of f1 and f2 increases. This results in reducing the overlapping of the frequency side bands.

This tuning - being saved as coefficients on a non-secure non volatile memory- could provide a backdoor to an attacker, because the higher frequency is shifted to the right and the lower one shifted to the left. In the case that an attacker clones the coefficients, the comparison sign bit could be reconstructed.

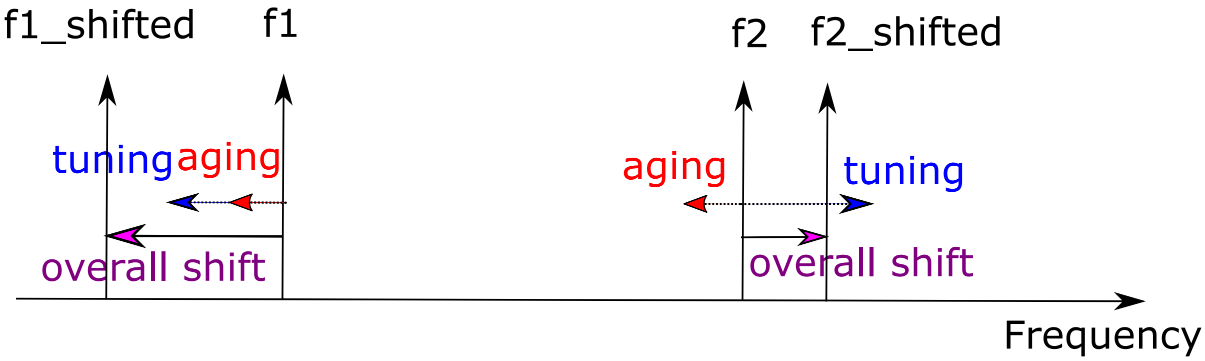


Figure 3.9: Tuning by shifting one frequency to right and one to left

3.4 Obtaining Tuning Coefficients, Simulation and Comparison of Results

The ring oscillator Physical Unclonable Function is designed in Cadence Virtuoso. Bit flip occurrences at the PUF output, are minimum for ring oscillator Physical Unclonable Function made out of whole ring oscillators with three inverter gates [41]. Hence ring oscillator units with three inverters are used.

Local variations were simulated by the Monte Carlo statistical analyses. $3sigma$ parameter range for oxide thickness T_{ox}, threshold voltage V_T, and gate length is taken into account. The standard deviation values come from simulation models, carried out for a 65nm CMOS technology.

Obtaining the Coefficients, and Performing Tuning Frequencies to Increase the Distance of Frequency expectations

Each oscillator's frequency gets varied with different variation mechanisms such as process, voltage, temperature, aging, and jitter.

$$f_{varied} = f_{original} + \Delta f_{process} + \Delta f_{temperature} + \Delta f_{Voltage} + \Delta f_{age} + \Delta f_{jitter} \quad (3.9)$$

$\Delta f_{process}$ defines the required margin due to process variation. $\Delta f_{temperature}$ defines required margin due to temperature corners. $\Delta f_{Voltage}$ defines required margin due to voltage corners. Δf_{age} defines required margin due to aging. Δf_{jitter} defines required margin due to jitter. In order to not have interference occurring in the frequencies of oscillator pairs, the side-band of each oscillator needs to be separated from another one.

$$With\ \Delta f = \Delta f_{process} + \Delta f_{temperature} + \Delta f_{Voltage} + \Delta f_{age} + \Delta f_{jitter} \tag{3.10}$$

$$iff1 < f2: \quad f2 > f1 + \Delta f \tag{3.11}$$

$$elseiff1 > f2: \quad f2 < f1 - \Delta f \tag{3.12}$$

where $f2$ is the frequency of second oscillator in pair, and $f1$ is the frequency of first oscillator in pair.

Therefore process, voltage, temperature, aging, and jitter analyses needs to be performed to determine the correct coefficients which prevent an interference to happen between the side-band of the oscillators in the pair, this adjustment is the meaning of what is referred to as "tuning". For obtaining the tuning coefficients, two sets of values need to be determined for the CAP_i and I_{INV} named as CAP_i^1, CAP_i^2 and I_{INV}^1 and I_{INV}^2. Based on formulas eq. (3.6) and eq. (3.7) this results in $CAP_{coefficient}^1$, $CAP_{coefficient}^2$, I_{INV}^1, I_{INV}^2. Indices 1 and 2 refer to oscillators 1 and 2 in the pair.

Tuning has been performed by taking the margins for paired sets of oscillators into account to meet the margins. It is not possible to solve the interference for every freely paired combination of oscillators in the population, hence based on the discussion in Section 3.2, a fully selected pairing approach is taken into account. This is an essential condition to meet the requirements driven through margins based on Equation (3.9) and Equation (3.11) and Equation (3.12). Although we are able to tune the selected pair perfectly with the needed margins in the Montecarlo sampling environment, it is not clear where the chosen samples out of Montecarlo end up post tapeout. Hence a process calibration on tester post tapeout is required to determine the process parameter and hence remove this uncertainty.

The tester based tuning is performed only on one corner i.e. the nominal VDD and the nominal temperature corner. We could take the following measures for taking the VDD, and the temperature corner effect into account.

1. The VDD corner:

   The ring structure is supplied through current sources instead of voltage sources, therefore the biggest part of the supply variation is naturalized through a fixed current source. In reality a very small variation might still remain after supplying the ring structure with the fixed current source supply. However we could take care of the remaining variation, as a pre-defined margin for the tester, to increase the difference between oscillator 1 and oscillator 2 frequencies based on this pre-defined margin.

   We defined this margin as $\Delta f_{voltage}$, and implemented it in the tuning. The exact value of the margin is determined using the Montecarlo samplings.

2. The temperature corner:

   The effect of the temperature corners on the inverters inside the ring structure remain and could not be neglected. This could not be naturalized through a fixed current source. However we could take care of the temperature variation as a pre-defined margin for the tester, to increase the difference between oscillator 1 and oscillator 2 frequencies, based on this pre-defined margin. We defined this margin as $\Delta f_{temperature}$, and implemented it in the tuning. The exact value of the margin is determined using the Montecarlo samplings.

3. The process corner:

   The process corner is determined by the tester through the single nominal

VDD/temperature corner test, and of course does not vary upon VDD and temperature corners.

As the process corner is determined and set after the tester tuning, the process uncertainty is removed therefore the equation is simplified to the following, excluding the process margin:

$$f2 = f1 + \ or \ \ -\Delta f \tag{3.13}$$

4. The aging: We defined this margin as Δf_{age}, and implement it in the tuning. The exact value of the margin is determined using the pre-tapeout aging simulations.

5. The jitter: We defined this margin as Δf_{jitter}, and implement it in the tuning. The exact value of the margin is determined using the pre-tapeout jitter simulations.

Montecarlo Sampling

Implementation of tuning adjustments is simulated in the Montecarlo sampling environment for a ROPUF with 10500 ring oscillator units (named as 10500 unit ROPUF for writing simplicity). 10500 is recommended by the Monte Carlo sampling environment as a suitable sample number being able to cover the process variation with 3 sigma range coverage. As no tester and tapeout was involved, the process was determined based on Montecarlo sampling. In order to analyze and substantiate the increment of expectation of frequency distance between the oscillators together with their desired margins, the tuning coefficients were designed with different minimum and maximum ranges of the distribution. The tuning A is designed to increase the minimum distance by a factor of 42, and the tuning B is designed to increase the minimum distance by a factor of 65. The exact value of this factor depends on the corners, noise conditions, and process. A factor of 42 is designed for less varied corner range:

$$vdd_{min} = 1V, vdd_{max} = 1.3V, temp_{min} = 0°C, temp_{max} = 80°C \tag{3.14}$$

$$aging\ margin = 3\%\ shift\ in\ f0 \tag{3.15}$$

Where f0 is the nominal oscillator frequency. Factor 65 is designed for a varied corner range:

$$vdd_{min} = 0.9V, vdd_{max} = 1.4V, temp_{min} = -40°C, temp_{max} = 125°C \tag{3.16}$$

$$aging\ margin = 5\%\ shift\ in\ f0 \tag{3.17}$$

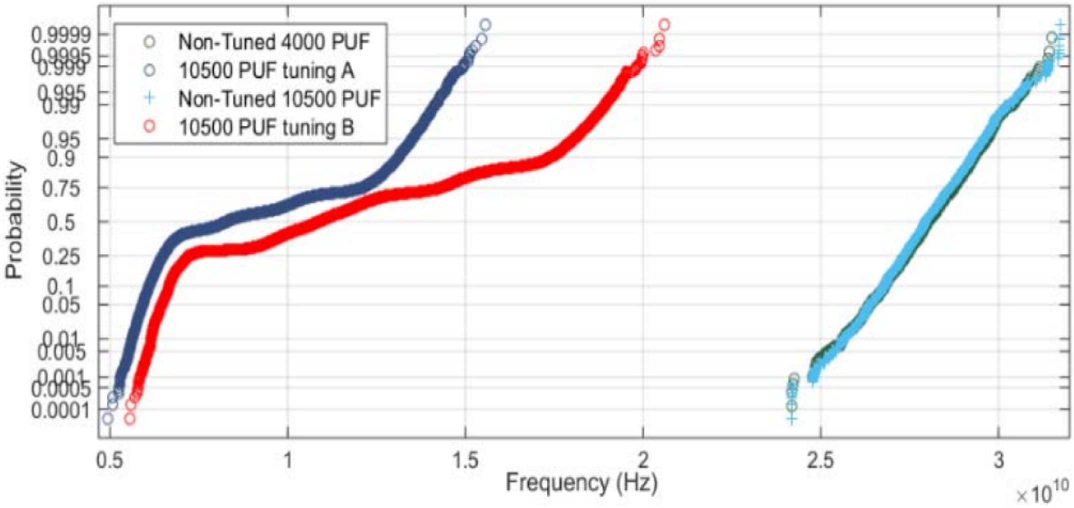


Figure 3.10: CDF Plots for the RO frequencies

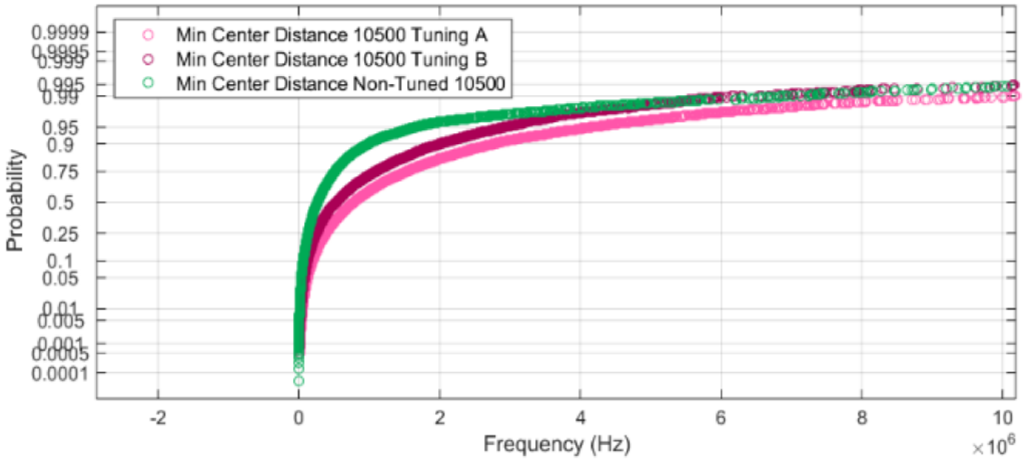


Figure 3.11: CDF plots for the RO side bands

Figure 3.10 shows the cumulative distribution function (CDF) for the tuned and non-tuned versions. The width of distribution has increased by the tuning, from $0.75e+10$ for the non-tuned ROs to $1.6e+10$ for tuning A showcasing an increment by 2.13 times. The width of distribution has also increased from $0.75e+10$ for the non-tuned ROs to $2.1e+10$ for tuning B showcasing and increment of 2.8 times. This results in reduced density of the RO frequencies. It could be observed that the mean frequency of the oscillators, are located at 28.0 $e+9$ for the non-tuned ones, and $8.0e+9$ for the A-tuned, and $11.2e+9$ for the B-tuned. These frequencies are higher than the in-field case due to lack of layout parasitics in the simulations. The tuning has reduced the frequencies due to the added load.

To get a closer look into the differences of frequency expectations, in Figure 3.11, distribution of frequency distances between two ROs in pair are plotted. The frequency distances are increased for tuned versions. The implemented margins in the tuned versions increase the reliability by reducing the overlaps. The tuning had increased the sideband distances.

There are two sets of tuning coefficients A and B applied in the Figure 3.10 and Figure 3.11. A difference between the tuning coefficients A and B is that, coefficient

B is set to result in a wider distribution of the frequency spectrum, resulting in a larger standard deviation. Also B is set to result in a higher maximum compared to A. Higher capacitance loads are set by tuning A to reduce the frequencies in a more dense manner compared to B. Tuning B hence results in a higher minimum distance between frequency side-bands and therefore better reliability than tuning A. The maximum and minimum of the distribution could be identified through the required interference resilience for side bands of the oscillators. This further could define the required minimum distance between oscillator frequencies.

3.5 Benefits and Limitations

Benefits

The proposed approach offers several notable benefits in enhancing the reliability of ring oscillator physical unclonable function against various sources of variation. It effectively addresses the challenges posed by aging, process variations, temperature fluctuations, voltage instabilities, and jitter. By considering and mitigating these factors, the reliability of ROPUF is significantly improved, ensuring more consistent and accurate performance.

Limitations

While the proposed approach brings valuable benefits, there are certain limitations that need to be acknowledged. To effectively design the coefficients and achieve practical implementation in the field, a tester-based determination of the process parameter is required. This adds complexity and may pose challenges for in-field coefficient design.

Furthermore, if the coefficients are stored on a potentially non-protected memory, the overall security of the approach may be compromised. In such cases, where third-party physical access to the device is possible, there is a risk of optical readout of on-chip memory values or probing of off-chip memory. To safeguard the memory values and ensure the integrity of the ROPUF, the use of protected fuses, or a protected

Non-Volatile Memory (NVM) becomes necessary. However, it should be noted that employing a protected-fuses or -NVM incurs additional costs and overhead.

In the case of an off-chip memory, the implementation of probing detection mechanisms can provide an added layer of protection. The forthcoming chapters in this work will delve into the subject of probing detection, exploring novel concepts and strategies to safeguard against potential security threats.

By identifying and discussing these limitations, the research aims to provide a comprehensive understanding of the benefits and challenges associated with the proposed approach. This enables informed decision-making and the development of countermeasures to address potential vulnerabilities, ultimately contributing to the advancement of ROPUF reliability and security.

3.6 Summary

In this body of work, a method is introduced to enhance the robustness of ring oscillator physical unclonable function designs. A theoretical model is formulated to guide the tuning of ring oscillators and improve their reliability against voltage, temperature, aging and other variation effects. By increasing the expected frequency differences between pairs of ring oscillators based on design margin requirements, as well as considering the impact of jitter, the proposed method achieves improved reliability against the variation effects. This novel approach stands apart from the state-of-the-art version, exhibiting greater resilience and distinguishable performance.

To ensure fairness and realism in the evaluation, the limitations of the proposed method are discussed. It is acknowledged that while this approach enhances reliability, it may be considered less secure when relying on a non-protected on-chip or off-chip memory. The vulnerability of the coefficients stored in an unprotected memory raises concerns about potential security breaches. However, in the case of an off-chip memory, protection against invasive microprobing can be employed to mitigate such risks. To address the issue of invasive microprobing, a novel probing detector concept will be presented in subsequent chapters, offering an additional layer of protection and security.

By acknowledging the limitations and focusing on addressing potential vulnerabilities, the research strives for a comprehensive and balanced exploration of the proposed method's strengths and weaknesses. This ensures a thorough understanding of its applicability and potential implications in real-world scenarios, leading to advancements in the field of ROPUF design and the overall security of hardware systems.

Chapter 4

Enhancing Security Measures Against Microprobing Attacks

4.1 Introduction

Microprobing, a sophisticated invasive attack technique, has gained significant recognition in various domains due to its potential implications for security. This method, as described in [47], is particularly noteworthy for its application in creating comprehensive firmware dumps for microcontrollers equipped with stringent security measures. The significance of this chapter lies in its extension of the author's previously published work, documented in [39]. The process of microprobing involves the careful attachment of a microprobe to the transmission lines responsible for transferring highly sensitive information, such as chip firmware or cryptographic keys. By establishing this connection, malicious actors gain the ability to extract valuable data from the targeted device, potentially compromising its integrity and security. To facilitate the attachment of the microprobe, researchers have explored the utilization of a cutting-edge technique known as Focused Ion Beam (FIB). This sophisticated method allows for precise editing of the chip's structural elements and can be employed from both the front and back sides of the chip. Such versatility and precision enhance the efficiency and efficacy of the microprobing process, further emphasizing its significance. The accessibility and relatively low cost of second-hand microprobing devices, as outlined in [32], underscore the importance of safeguarding against microprobing attacks. While high-security devices are undoubtedly a prime concern, it is crucial to recognize that the threat extends beyond such specialized contexts. Therefore, implementing protective measures against microprobing becomes imperative across a wide range of devices and applications, where sensitive information is at risk.

Considering the potential ramifications of microprobing attacks, it is paramount to devise comprehensive strategies and countermeasures. These measures should address the vulnerabilities exposed by the ease of acquiring second-hand microprobing devices and ensure the protection of sensitive data. By proactively implementing robust security measures, organizations can mitigate the risks associated with microprobing attacks and safeguard their systems against unauthorized access and potential compromise.

One effective countermeasure against probing attacks involves the implementation of bus encryption and masking techniques to conceal sensitive information [23]. However, it is important to note that these approaches impose considerable processing overhead, leading to substantial data transfer delays across the data bus [51]. While they offer enhanced security, the reduced data throughput adversely affects the efficiency of bus communication.

An alternative approach is to employ physical obstruction-based solutions. These solutions aim to prevent unauthorized physical access to the metal nets responsible for carrying critical data by employing either active or passive meshes on the top metal layer. Passive meshes serve to shield and render the lines of interest more challenging to access. However, it is crucial to acknowledge that skilled attackers may still find ways to bypass these passive meshes by milling through the mesh and gaining access to the covered lower layers [49].

In contrast, active meshes placed on top of the protected lines can actively detect probing attempts by continuously monitoring the integrity of the mesh. By actively measuring whether the mesh has been compromised, these active meshes serve as an additional layer of defense. However, it is important to note that both active and passive mesh solutions necessitate the inclusion of an extra metal layer dedicated solely to the protective mesh, thereby rendering them cost-intensive approaches. Furthermore, it should be emphasized that these solutions do not offer protection against backside probing, which poses an additional security concern [51].

Another potential approach to safeguard against probing attacks involves the utilization of a tamper protection foil, as proposed by previous research [22]. However, it is crucial to highlight that this foil is not integrated directly on the chip, necessitating the incorporation of supplementary circuitry for measurement purposes, along with the physical envelope encompassing the circuit board. While this approach offers an additional layer of protection, it introduces additional complexities and considerations in terms of design and implementation.

An intriguing alternative approach involves the utilization of *analog sensors* to detect the effects of microprobing on the primary circuitry, as proposed by various researchers [49]. The Probe Attempt Detector (PAD) [34], for instance, employs analog elements such as on-chip capacitance and current sources to discharge a capacitor based on the delay discrepancy between a probed and unprobed line. By leveraging this analog sensing mechanism, the PAD effectively protects the chip against both front and backside probing, all while mitigating the latency associated with data encryption. However, it is essential to acknowledge that the inclusion of on-chip capacitance presents challenges in terms of area consumption and cost. Additionally, the availability of the required capacitance may vary depending on the specific manufacturing processes [51]. Furthermore, it is worth noting that the utilization of analog elements necessitates a time-consuming analog design flow, resulting in increased design complexity and extended time-to-market.

More recently, *digital sensors* have emerged as an alternative to reduce the complexity and overhead associated with analog design. Weiner *et al.* [52] proposed a comprehensive digital solution known as the *Low Area Probing Detector (LAPD)*, which relies on a race condition to compare the delay between two bus lines. However, it is important to consider that achieving high yield and security with the LAPD requires meticulous balancing of the protected lines and fine-tuning of the LAPD transistor dimensions [19]. To address these challenges, Weiner *et al.* introduced

CaLIAD [51], a novel approach that also leverages a race condition on two bus lines. Unlike the LAPD, CaLIAD incorporates a Vernier Delay Line Time to

Digital Converter (VDL-TDC) to detect the probe effect on the bus line delay. Nonetheless, the utilization of a VDL-TDC introduces additional complexities, particularly concerning local variations. Firstly, to detect microprobes accurately, a sub-gate-delay picosecond resolution TDC is required, which inherently exhibits sensitivity to mismatch [51]. Secondly, since a VDL-TDC employs two non-symmetric delay lines, it fails to naturally nullify the dual rail variation through rail differentiation, thus necessitating additional considerations. Consequently, the practical implementation of CaLIAD necessitates extensive post-production characterization efforts, including the generation of individual calibration values for each pair of bus lines on each chip. These calibration values must be stored in a secure manner to prevent manipulation, adding to the test effort and protected memory overhead. Additionally, to ensure high reliability, the probed state must be emulated during calibration by attaching a capacitance equivalent to the smallest probe to be detected to the bus lines. This requirement introduces on-chip capacitance, further contributing to the overall overhead. Furthermore, it is crucial to recognize that the calibration process is dependent on the test capacitance, and any deviations resulting from process variations or intentional modifications by malicious actors may weaken the detection sensitivity [51].

In this work, we present the fully digital and highly predictive Ring Oscillator Probing Attempt Detector (ROPAD), which offers significant advancements over the state of the art in various aspects. We compare ROPAD with previous works to highlight its advantages:

Protective Mesh: ROPAD provides protection against backside attacks without the need for an additional metal layer, eliminating any overhead associated with it.

Bus Encryption: Unlike existing techniques that require processing effort and introduce latency for encryption, ROPAD offers protection without such overhead.

PAD [34]: ROPAD eliminates the need for an analog design flow, simplifying the implementation process. It also avoids the use of expensive on-chip capacitors.

LAPD [52]: ROPAD reduces the calibration effort and associated overhead compared to LAPD. It eliminates the need for transistor fine-tuning and provides a better detection margin and yield. Moreover, ROPAD offers protection for bus capacitance four times higher than LAPD.

CaLIAD [51]: ROPAD minimizes the calibration effort and overhead required compared to CaLIAD. It also reduces the amount of protected memory needed. Similar to LAPD, ROPAD provides protection for bus capacitance four times higher than CaLIAD.

Our proposed concept not only reduces calibration overhead and required protected memory but also enables the protection of longer buses with higher capacitance. Additionally, we consider the protection of cross-coupled bus lines, which has not been addressed in previous works.

The key contributions of our work are as follows:

1. We introduce ROPAD, a novel oscillator-based Probe Attempt Detector. The concept is theoretically motivated at the design level. ROPAD enables the detection of a 20fF probe, as desired according to [51], on a bus with four times higher intrinsic capacitance compared to current approaches. It offers either no or low on-chip calibration overhead and reduces the number of associated protected data by a factor of $2N$. The symmetric design of our concept ensures the highest resilience against local process variations.

2. We provide a comprehensive simulative evaluation of our new Probe Attempt Detector in a sub-40nm[1] technology, considering local process variations through Monte Carlo analysis. The evaluation takes into account various factors such as process variations, ensuring a robust assessment of the detector's performance under realistic conditions.

3. Furthermore, we conduct a thorough simulative corner analysis of the Probing Attempt Detector, covering a range of corner cases including fast NMOS fast PMOS (FNFP), slow NMOS slow PMOS (SNSP), FNSP, and SNFP corners,

[1] The specific technology is not disclosed due to confidentiality reasons.

as well as variations in temperature and supply voltage. This analysis provides valuable insights into the detector's behavior and performance across different operating conditions.

4. In addition to the aforementioned contributions, we are the first to consider the impact of cross-coupling between bus lines during probing. By investigating this aspect, we uncover and demonstrate the effect of cross-coupling, which has been overlooked in previous studies. This finding expands our understanding of probing attacks and contributes to the development of more robust countermeasures.

These contributions collectively advance the field of probing attack detection and mitigation, providing novel insights and solutions that enhance the security of integrated circuits.

4.2 Detection for Non-Coupled Bus Lines

The probing detection core attached to a regular bus as the device under protection is depicted in Figure 4.1.

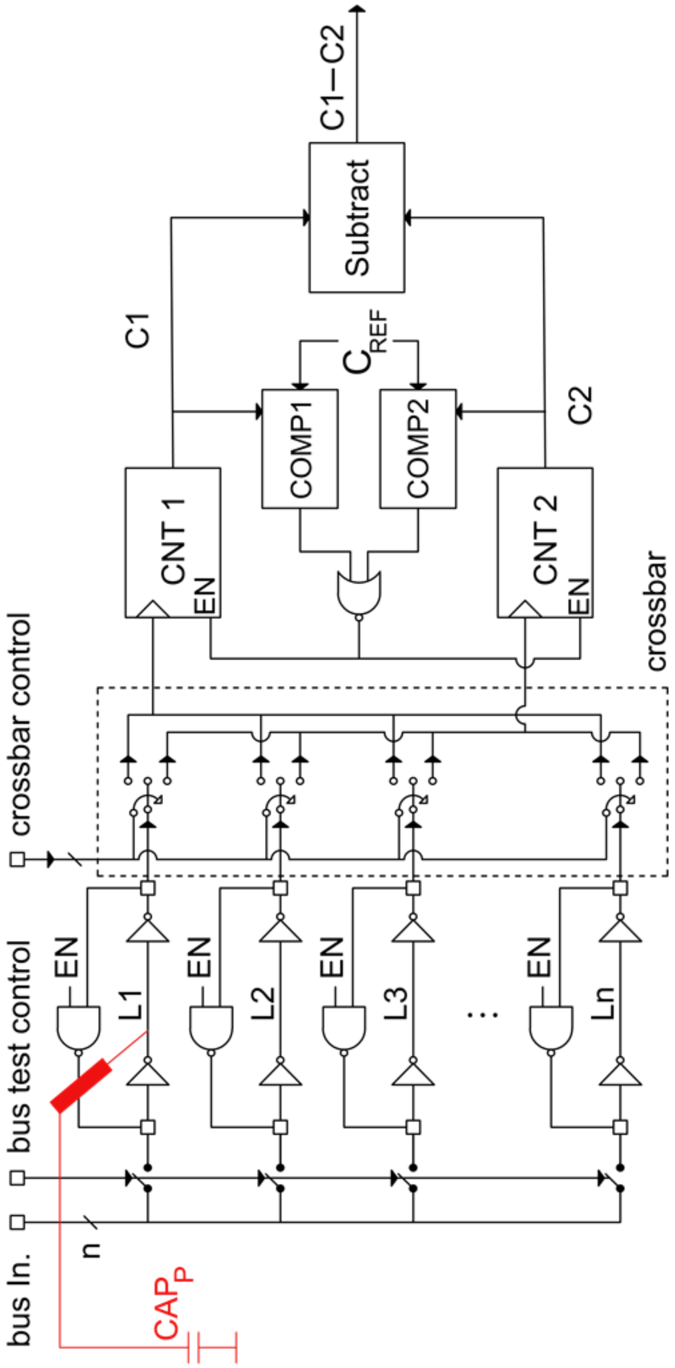


Figure 4.1: System view of the regular data bus probing detection concept ROPAD.

On the left an N-line data bus is shown containing lines L1 to Ln. Each line is buffering the data using four inverter gates. In the normal mode of the data bus, it is loaded with the input data from *"bus in."* In the test mode, the bus test control unloads the bus lines from the input data through its controlled switches. Moreover the outputs of the bus are unloaded from the normal circuit operation through the *cross-bar*. Hence the bus is in test mode.

Each bus line usually consists some buffer/inverter cells to buffer the data through the metal routing of that line during its normal operation. To save area, these cells are reused in our concept to make the bus lines probing-aware. In the test mode, a feedback is established on the bus through an additional NAND gate. When an enable signal is set at the second NAND gate input, this results in a loop of an odd number of inverters, i.e. a ring oscillator (RO). The period of the RO's oscillation directly depends on the propagation delay of the physical bus line and its loading elements. It is therefore sensitive against probing.

ROPAD Model for Non-Coupled Lines

In the case where a probe is placed on a bus line, the capacitance of the probe adds a load to the line, thereby increasing its propagation delay and the oscillation period of the formed Ring Oscillator (RO). The capacitances of unprobed and probed lines are given by:

$$CAP_{un} = CAP_L, \tag{4.1}$$

$$CAP_{pr} = CAP_L + CAP_P, \tag{4.2}$$

Here, CAP_P represents the probe capacitance, and CAP_L corresponds to the total bus line capacitance. For non-coupled bus lines, the frequency of the RO can be expressed as:

$$f_L = \frac{1}{2\sum_{i=1}^{M} d_i}, \tag{4.3}$$

where M denotes the number of inverters in the line, and d_i represents the delay per stage. The estimation of d_i based on the *alpha-power* model for transistors is given by [43, 12]:

$$d_i = k \cdot \frac{CAP_i \cdot V_{DD}}{(V_{DD} - V_T)^{\alpha}}, \tag{4.4}$$

In this equation, V_{DD} represents the supply voltage of the bus rails, CAP_i denotes the output capacitive load of an inverter gate, V_T represents the threshold voltage, α corresponds to the velocity saturation coefficient of the carriers, and k represents the trans-resistance [8]. Introducing a technology factor φ defined as:

$$\varphi = k \cdot \frac{V_{DD}}{(V_{DD} - V_T)^{\alpha}}, \tag{4.5}$$

the frequency f_L of the RO in eq. (4.3) can be expressed as:

$$f_L = \frac{1}{2\varphi \cdot \sum_{i=1}^{M} CAP_i}. \tag{4.6}$$

For the initial model, we assume that all bus lines have the same delay. Consequently, the approximate frequency difference between any unprobed and probed line can be calculated as follows:

$$\Delta f = f_{L,un} - f_{L,pr} = \frac{1}{2\varphi}\left(\frac{1}{\sum_{i=1}^{M} CAP_{L,i}} - \frac{1}{\sum_{i=1}^{M}(CAP_{L,i} + CAP_{P,i})}\right), \tag{4.7}$$

where $f_{L,un}$ represents the RO frequency for an unprobed line, and $f_{L,pr}$ represents the RO frequency for a probed line.

To detect the presence of a probe, we measure the frequency difference between two such ROs using two counters, namely *CNT1* and *CNT2*, which count the oscillations. Comparators *COMP1* and *COMP2* compare the counter values to a reference value C_{REF}. The first counter to reach C_{REF} triggers an interrupt, disabling the counters by setting EN to logical 0. We name this form of connecting the counters as "cross coupled counters". In the case where $f_{L,un}$ is equal for both bus lines ($f_{L,un} > f_{L,pr}$), the counter value for the unprobed line becomes $C_{un} = C_{REF}$, and the counter value for the probed line becomes $C_{pr} < C_{un}$. Using eq. (4.7), we can obtain the resulting values for C_{un} and C_{pr}:

$$C_{un} = C_{REF}, \tag{4.8}$$

$$C_{pr} = C_{REF} \cdot \left(1 - \frac{\Delta f}{f_{L,un}}\right). \tag{4.9}$$

The expected output of the cross-coupled counters, denoted as $(C_1 - C_2)$, where $C_1 = C_{un}$ and $C_2 = C_{pr}$, is given by:

$$\Delta C(C1, C2, C_{REF}) = (C_{un} - C_{pr}) = C_{REF} \cdot \left(\frac{\Delta f}{f_{L,un}}\right). \tag{4.10}$$

The term C_{REF} denotes the fact that the difference of counters are given the cross coupled form of the counters, based on a reference value C_{REF}. In this thesis from now on, we use $C1 - C2$ for simplification purposes, instead of using the term $\Delta C(C1, C2, C_{REF})$.

Based on which counter - C1 or C2 - wins the race to reach the reference value, this function could be defined in more detail as:

$$\Delta C(C1, C2, C_{REF}) = \begin{cases} C_{REF} - C2, & \text{for } C_1 = C_{REF} > C2 \\ C1 - C_{REF}, & \text{for } C_2 = C_{REF} > C1 \end{cases}. \tag{4.11}$$

By employing the counters in a differential form, we gain two main advantages:

1. There is no need for a reference counter that relies on the system clock, thus

mitigating potential attacks on the reference and removing the effects of variations or drift in the reference frequency.

2. Differential measurement partially compensates for environmental effects and global process variations.

To minimize the number of counters, the main probing detection core can share two counters, which are reset before each test, among all bus lines. By employing N non-shared coupled counter pairs, the test duration is reduced, albeit with a corresponding increase in area overhead.

System Integration of the Probing Detection Core

The system integration of the probing detection core involves comparing Ring Oscillators (ROs) formed from different bus lines and extracting the counter difference. However, probing adjacent lines presents an unstable setup, making it increasingly unlikely to successfully build a probing setup as the number of probes increases [51].

Under the assumption that at least one bus line remains unprobed, we propose a system integration principle, as illustrated in Figure 4.2, to protect an N-line bus with $N-1$ probes attached. Each bus line is compared with two other lines in an overlapping manner. For instance, Line 1 (L1) is compared to Line 2 (L2), L2 to L3, and so on, until the last line (LN) is compared to L1.

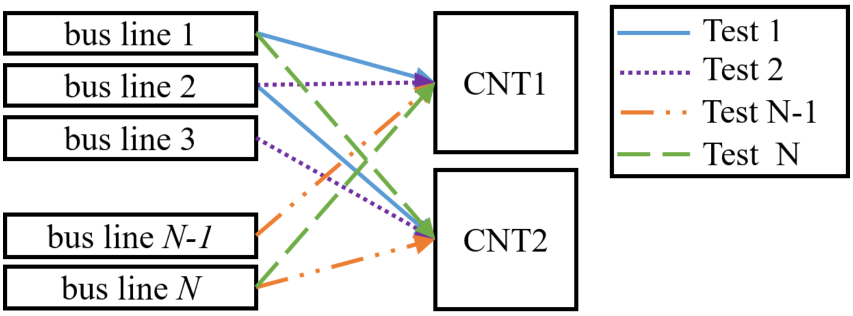


Figure 4.2: Visualization of test sequence

By setting the alarm threshold as $C_1 - C_2 \geq DB$, the case where the line (L1) connected to *CTR1* is unprobed, and the line (L2) connected to *CTR2* is probed can be detected. This case is referred to as *forward probing*.

In our test scheme, each line is connected once to *CTR1* and once to *CTR2*. However, the case where L1 is probed and L2 is unprobed, known as *reverse probing*, remains undetected in the test. Nevertheless, since we assume that at least one bus line is unprobed, there will be at least one test where an unprobed line is connected to *CTR1* and the probed line is connected to *CTR2*. Therefore, considering the forward probing case is sufficient, while the mention of reverse probing is solely for completeness.

Corner Analysis for the Non-coupled Bus Lines

The probing detection core and the bus lines have been implemented using cell models in a CMOS sub-40nm technology. In this section, we focus on the analysis of the non-coupled bus lines, which provides a simpler case for examination.

To assess the detector's performance under global process variations as well as variations in voltage and temperature (PVT), we conducted a corner analysis. This analysis was conducted for different process corners, including Fast Nmos Fast Pmos (FNFP), Slow Nmos Slow Pmos (SNSP), Fast Nmos Slow Pmos (FNSP), and Slow Nmos Fast Pmos (SNFP), as well as various voltage and temperature settings within the ranges of $V_{DD,max}$, $V_{DD,min}$, T_{max}, and T_{min}.

The corresponding corner analysis data for the counter difference is depicted in Figure 4.3. The upper surface, colored in yellow, represents the probed case, while the middle surface, colored in green, represents the unprobed case. The lower surface, colored in blue, represents a reversed probed case where the probe is placed on Line 2 instead of Line 1. The inclusion of the reversed probed corner analysis is important to consider potential scenarios where an attacker attempts to bypass the detector by placing the probe on the other line. However, the test sequence implemented in our approach can detect the reversed probe case. The results of the reversed probed corner analysis are included here for the sake of completeness.

In this analysis, the worst-case corners are denoted as X, Y, and Z. The SNSP corner yields the minimum value (worst case) for the counter difference when one line is probed and the other line remains unprobed. The slow Pmos and Nmos in the inverter gates contribute to a reduced propagation delay and, consequently, a reduced oscillation frequency. This reduction is significant for both counters, resulting in the worst case process corner. The ultimate worst-case corners occur when both the VDD and temperature reach their minimum values, and the process corner is SNSP. It can be observed from Figure 4.3 that these worst-case corners, represented by points X, Y, and Z, exhibit the smallest $C1 - C2$ values. Therefore, distinguishing a probe becomes more challenging for these worst-case corners, as the decision boundary must be placed within a narrower numerical range. In the subsequent steps, we will explore how considering the statistical distribution of process samples affects this smaller numerical range and its impact on the sensitivity and yield of the detector. Based on these results, we will further characterize ROPAD for the worst corners X, Y, and Z using statistical, Monte Carlo-based mismatch analysis.

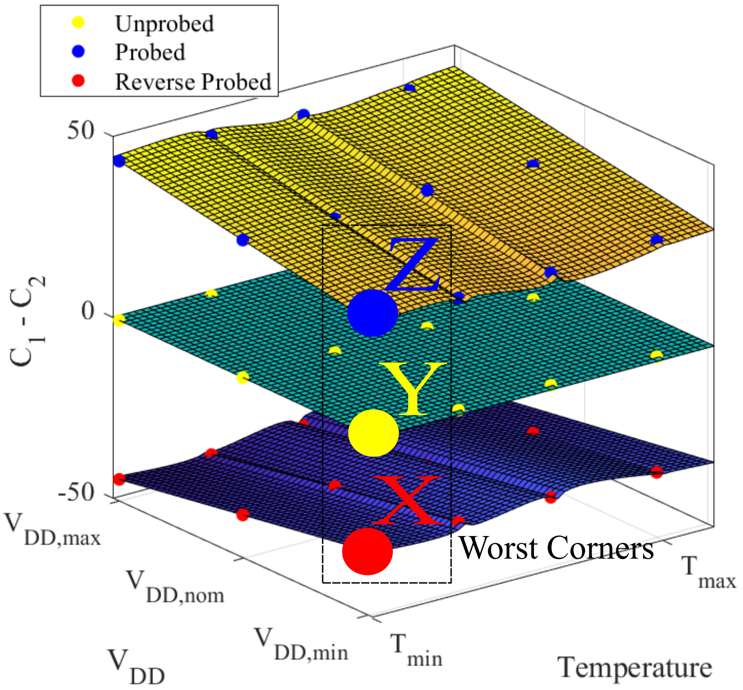


Figure 4.3: Corner analysis over temperature and voltage for SNSP with surfaces fitted to simulated points assuming $CAP_P = 20fF$, $CAP_L = 400fF$. X, Y, Z are worst case corners.

4.3 Statistical Analysis with the Non-Coupled Bus Lines

In this section, we delve into the statistical distribution analysis of the non-coupled bus lines. While the ideal scenario assumes perfect equality and no mismatch, any deviation from $C_1 - C_2 = 0$ indicates the presence of a probe. However, in real-world scenarios, local mismatches among the observed lines lead to a statistical distribution of counter differences.

To analyze this effect, we conducted a Monte Carlo mismatch analysis for the worst-case corners X, Y, and Z, as described in Section 4.3. We collected 2000 samples considering a local variation of the process within $4.5 \cdot \sigma$. This analysis was performed using an analog circuit simulator from Cadence. We modeled the results using Gaussian distributions with mean μ_u and variance σ_u for an unprobed pair of bus lines (curve Y), and mean μ_p and variance σ_p for a forward probed pair of bus lines with probe capacitance $CAP_P = 20fF$ (curve Z). Please note that the exact data points cannot be published due to FAB non-disclosure agreements. The resulting fitted cumulative distribution is presented in Figure 4.4.

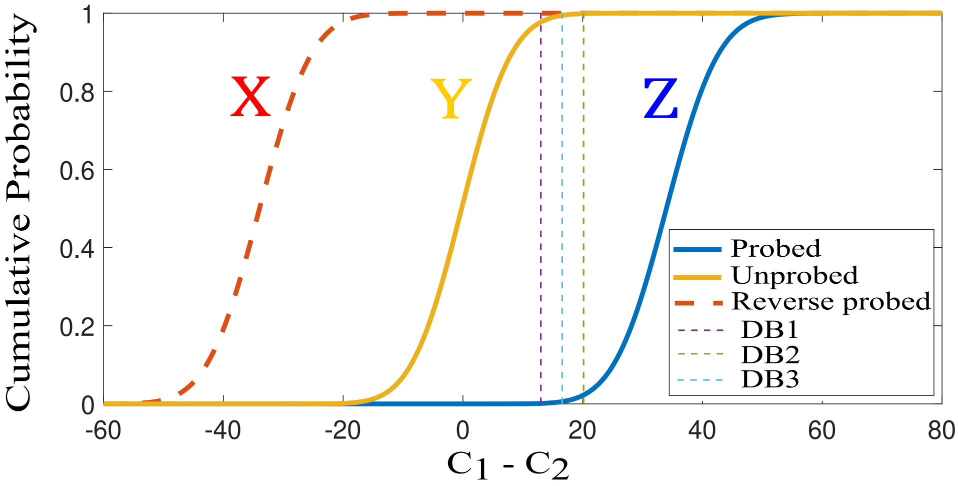


Figure 4.4: Gaussian fitted $cdf(C_1 - C_2)$ at worst-case corners X, Y, Z.

As both the probed and unprobed states exhibit distributed counter values, distinguishing between the two cases becomes a probabilistic task. To address this challenge, we introduce the concept of Decision Boundaries (DBs) for probe detection. The generation of a probing alarm is based on a specified boundary, as defined by the following function:

$$\text{Alarm Boolean} = \begin{cases} \text{false}, & \text{for } C_1 - C_2 < DB \\ \text{true}, & \text{for } C_1 - C_2 \geq DB \end{cases}. \tag{4.12}$$

In other words, an alarm is triggered when the counter difference exceeds the designated DB. For further analyses, we define three DBs and evaluate ROPAD's resulting yield and detection sensitivity, as discussed in Section 4.4.

The first DB, denoted as $DB1$, is defined as $\mu_u + k\sigma_u$, where k represents the number of standard deviations (σ) of the unprobed state. This boundary aims to separate the $CAP_P = 20fF$ probed state from the unprobed state by $k\sigma_u$.

The second DB, denoted as $DB2$, is defined as $\mu_p - k\sigma_p$, representing $-k\sigma_p$ of the $CAP_P = 20fF$ probed state. This boundary is designed to achieve a $k\sigma$ sensitivity goal.

Finally, $DB3$ represents a trade-off between sensitivity and yield. It is computed such that

$$1 - cdf_u(DB3) = cdf_p(DB3), \tag{4.13}$$

where $cdf_u(C_1 - C_2)$ and $cdf_p(C_1 - C_2)$ are the resulting cumulative distribution functions for the unprobed and probed states, respectively.

The resulting DBs (for $k = 2$) are also depicted in Figure 4.4. In the experiment shown, the probed case (curve Z) is distinguishable from the unprobed state (curve Y) using all the DBs. However, the reversed probed case (curve X) cannot be detected, as discussed in Section 4.2. It is worth noting that our analysis includes not only V_{DD} and temperature variations but also global process variations. Therefore, the defined DBs are applicable to all chips without the need for individual calibration. The chosen DB can be hard-coded, for example, in a read-only memory (ROM).

4.4 Figures of Merit

In the analysis of ROPAD, we introduce two important metrics: yield (yld) and sensitivity ($sens$). These metrics have a maximum value of one, representing optimal performance.

A yield value of $yld = 1$ implies that the ROPAD system does not trigger an alarm on any device under worst-case conditions when no probe is present. This corresponds to a zero False Positive rate (FP).

On the other hand, a sensitivity value of $sens = 1$ means that the system always triggers an alarm when a probe with a capacitance of $CAP_P \geq 20fF$ is placed. This corresponds to a zero False Negative rate (FN).

The metrics FP, FN, yld, and $sens$ are defined as follows:

$$FP = 1 - cdf_u(DB) \tag{4.14}$$

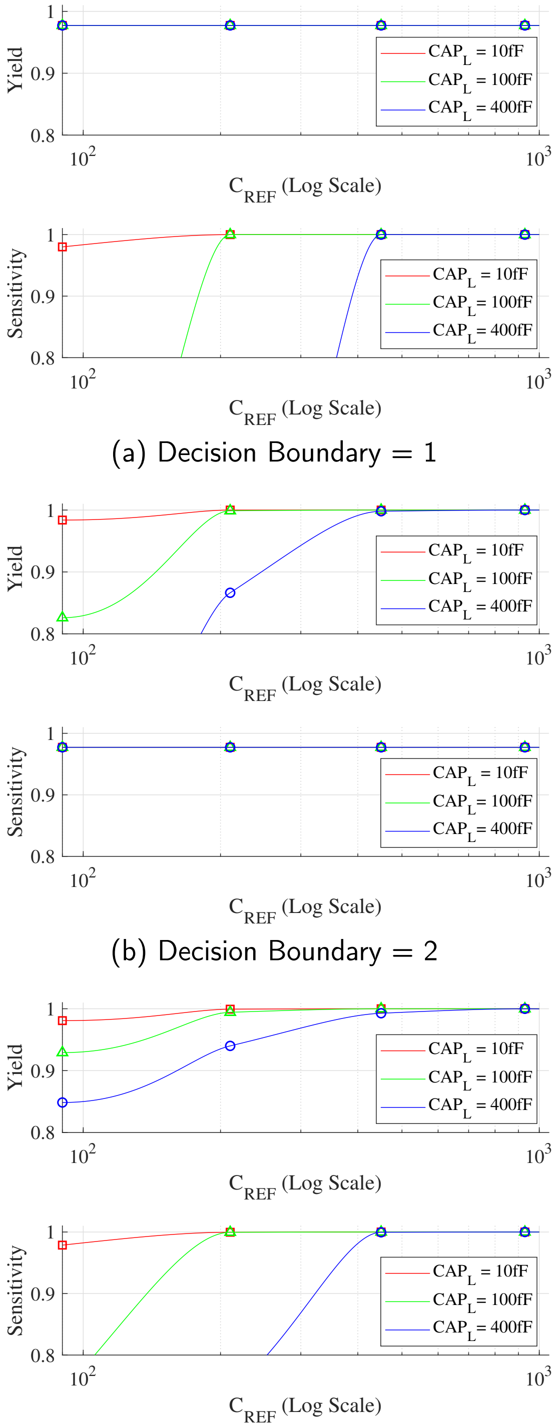
$$yld = 1 - FP = cdf_u(DB) \tag{4.15}$$

$$FN = cdf_p(DB) \tag{4.16}$$

$$sens = 1 - FN = 1 - cdf_p(DB) \tag{4.17}$$

Here, DB represents the Decision Boundary.

In Figure 4.5, we plot the values of yld and $sens$ for $DB1$, $DB2$, and $DB3$ as functions of C_{REF} for three different intrinsic line capacitances of non-coupled lines, assuming a probe capacitance of $20fF$.



(a) Decision Boundary = 1

(b) Decision Boundary = 2

(c) Decision Boundary = 3

Figure 4.5: Yield and sensitivity, plotted as a function of C_{REF} assuming that a $20fF$ probe is placed on different length lines.

Based on previous studies [51, 52], it is known that increasing the probe capacitance does not significantly affect the yield but makes it easier to detect. Therefore, testing for the worst-case scenario of the smallest existing $20fF$ active probe [51] is sufficient. From Figure 4.5, we observe that the chosen value of $k = 2$ results in $yld = 97.9\%$ for $DB1$ and $sens = 97.9\%$ for $DB2$ across all values of C_{REF}.

Furthermore, both $sens$ and yld converge towards 100% for $DB1$ and $DB3$, as well as for $DB2$ and $DB3$, as C_{REF} (i.e., the test time) increases. It is expected that the convergence is slower as the metrics approach 100%.

The results validate that in ROPAD, the test time can be traded for yield and sensitivity by doubling C_{REF}, which corresponds to doubling the test time. Hence, no calibration effort is required for ROPAD in the case of non-coupled bus lines, as long as the values of C_{REF} and DB are carefully selected during the design phase.

However, it is worth noting that post-silicon calibration, although incurring calibration overhead, can be used to further improve the yield and sensitivity of ROPAD. Unlike CaLIAD [51], which requires characterization under emulation of an attack and an on-chip test capacitance, the calibration of ROPAD determines the realization of the distribution in Figure 4.4, and a global DB is adjusted accordingly to enhance yld and $sens$.

When designing ROPAD, we recommend selecting either $DB1$, $DB2$, or $DB3$ based on specific design requirements, with $DB3$ being the most suitable choice in most cases. Additionally, adjusting C_{REF} to achieve the desired yield and sensitivity is advisable as long as the runtime remains acceptable. Per-chip calibration is only suggested if the runtime of ROPAD is otherwise too high.

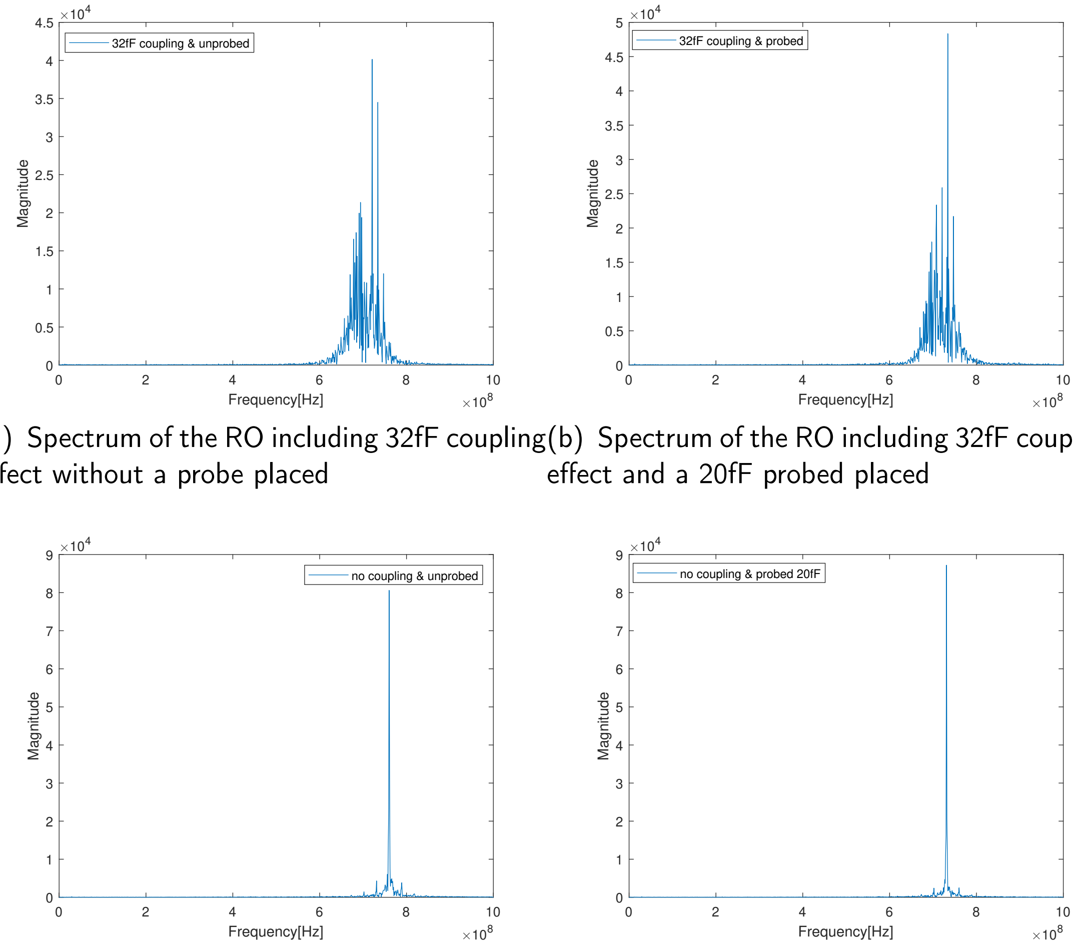
Pre-silicon determination or in-silicon calibration of line-pairs is necessary only if the assumption of equally long bus lines is violated.

4.5 Detection for Coupled Bus Lines

Capacitive coupling between the bus lines results in precharging one bus line through transitions at the other bus line.

As a consequence the propagation delay of the transitions in probed and unprobed line get closer and the two states are harder to distinguish. We have extracted the RO waveform out of the simulations to Matlab and plotted the spectrum of the RO in the frequency domain using the Matlab scripting. To see the effect of coupling and the probe, We have performed this for several cases including the coupling capacitance, excluding the coupling capacitance, with a probe and without a probe placed on the lines. In the following we showcase, and discuss the results. Figure 4.6 shows the effect of coupling capacitance on the ring oscillator output spectrum, by comparing 4.6a / 4.6b with 4.6c / 4.6d we can see that the coupling flattens the spectrum and increases the bandwidth. Figure 4.6 also shows the effect of placing a microprobe on the RO spectrum, comparing the fig. 4.6c and the fig. 4.6d we can see that the probe shifts the peak to the left due to its induced capacitive load. The ones including no coupling (4.6c and 4.6d) show a clear peak behavior and its shift, however the ones including coupling effect (4.6a and 4.6b) show a rather flattened spectrum behavior in which a clear shift of the probe peak is less observable, but rather a total shift of the bandwidth to the left could be observed, in which the shift is from a smaller order than the overall bandwidth. The spectrum of the ROs with and without 32fF coupling effect, shows the effect of 32fF coupling in flattening the spectrum (comparing 4.6a and 4.6c). Spectrum of the ROs including a 20fF probe, with no coupling effect taken into account, shows the effect of the probe on shifting the spectrum (comparing 4.6c and 4.6d).

We could observe that by introducing the coupling, the oscillation is distributed around the probe peak. As a result this makes the RO output less stable. The probe does not flatten the spectrum, but only shifts the peak to the left, by a frequency difference which is smaller than the bandwidth of the coupled unprobed RO. This, of course, makes the probe detection less successful, due to the fact that the probe effect is smaller order than of the coupling.



(a) Spectrum of the RO including 32fF coupling effect without a probe placed

(b) Spectrum of the RO including 32fF coupling effect and a 20fF probed placed

(c) Spectrum of the ROs with no coupling effect taken into account, without a probe placed

(d) Spectrum of the ROs including a 20fF probe, with no coupling effect taken into account

Figure 4.6: spectrum of the ROs (first harmonic)

Line Model for Capacitively Coupled Bus Lines

In order to accurately model the delay of capacitively cross-coupled bus lines, the previously used *alpha-power* model, as given by Equation 4.4, is no longer sufficient. This is primarily due to its lack of consideration for transition dependency.

To address this limitation, full transient SPICE-based simulations with a parasitic extracted line model are employed, providing a more precise approximation.

The resulting line model for the inverter-to-inverter cross-coupled line parasitics is illustrated in Figure 4.7. It incorporates discrete resistors $R1$ and $R2$, representing the bus line resistances for Line 1 and Line 2, respectively. Additionally, the model includes discrete capacitors CAP_G to account for the line-to-ground capacitance values from the layout. The coupling capacitances between the bus lines are represented by discrete capacitive elements CAP_C. The equivalent circuit diagram of a bus line is shown in Figure 4.7, considering three discrete elements. For a higher level of accuracy, 20 discrete elements per line between each pair of inverters are utilized in the actual evaluation.

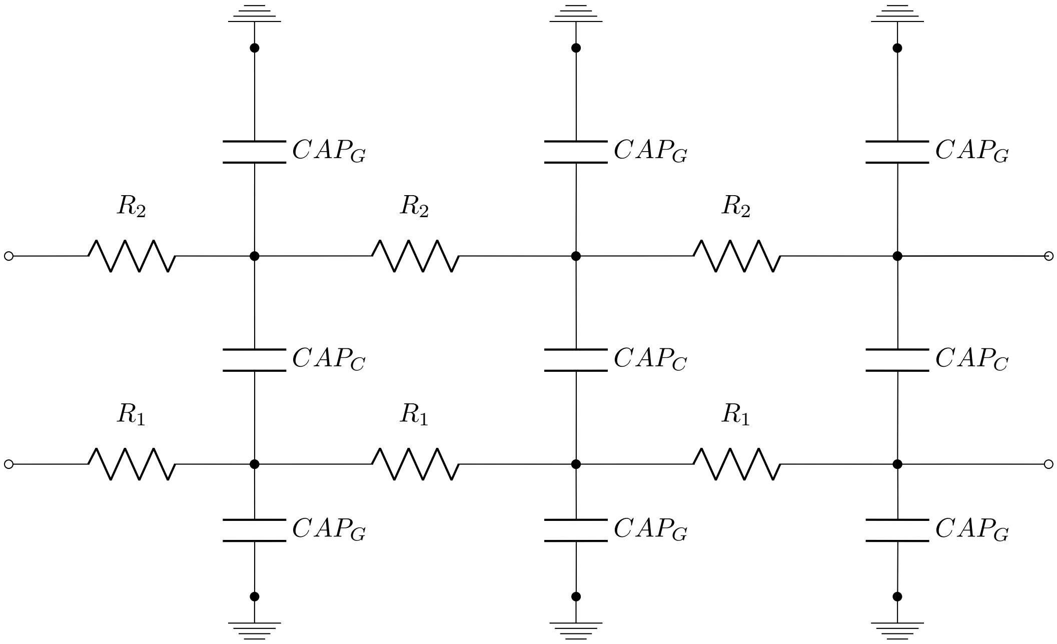
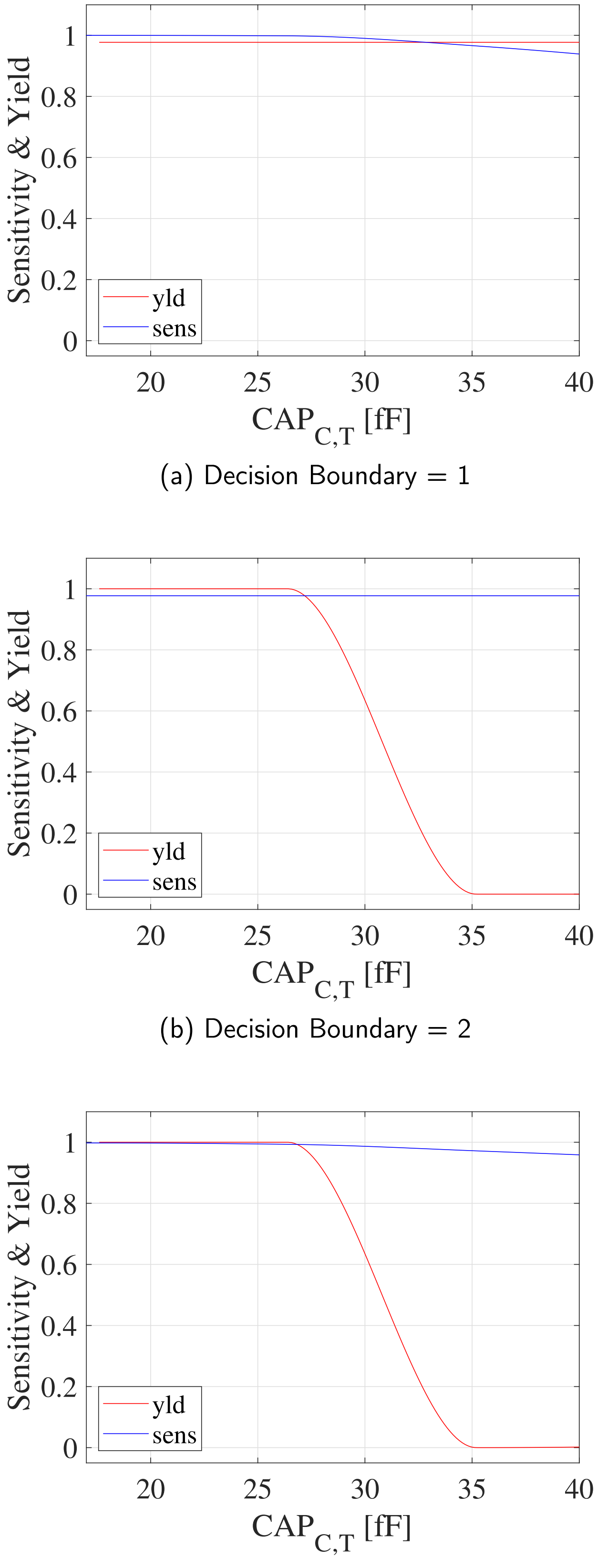


Figure 4.7: Simulation model for coupled line segments between each two inverter stages in the bus.

Experiments

For our experiments, the parasitic values including the total coupling capacitance for the bus line - named as ($CAP_{C,T}$) - for a $400fF$ data bus are extracted. The $CAP_{C,T}$ value depends on the bus line separation. Therefore, $CAP_{C,T}$ is sweeped between $17fF$ and $160fF$. *sens* and *yld* of ROPAD are again approximated for each case by Monte Carlo analyses.

Results using the DBs from above with $k = 2$ are shown in Figure 4.8.



(a) Decision Boundary = 1

(b) Decision Boundary = 2

(c) Decision Boundary = 3

Figure 4.8: Yield and sensitivity plotted as a function of the coupling capacitance for a $400fF$ bus line, assuming that a $20fF$ probe is placed, and $C_{REF} = 500$.

Out of the measurements, we could understand that $sens$ is reduced to less than 97.9% (or 3 sigmas CDF) at $CAP_{C,T} > 35.6fF$, for $DB1$ and $DB3$. The $CAP_{C,T} > 35.6fF$ is basically the highest coupling capacitance that this version of the detection core can tolerate for establishing the sensitivity, given that $DB1$ and $DB3$ as decision boundaries. So this value is just an observation out of measurements, and is not representing a pre-designed factor. As it were discussed earlier in this chapter, $DB1$ is designed to deliver 97.9% yield (or 3 sigmas CDF). Also $DB2$ is designed to deliver 97.9% (or 3 sigmas CDF) sensitivity. yld and $sens$ close to 100% are only achieved for DB3 and until $CAP_{C,T} = 25fF$, as it is obtained per measurements. This shows the significance of the coupling effect which was yet unconsidered in PAD constructions.

Updating the test sequence so that neighboring bus lines are not considered in the same test, but always at least one bus line in between the tested ones is grounded, realizes a shield between the lines and might solve the problem not only for ROPAD but also for other PADs. However, at least a reduction to $CAP_{C,T} < 25fF$ can be expected, which suffices for ROPAD with $C_{REF} = 500$. Alternatively, for ROPAD increasing C_{REF}, i.e. the test time, would increase yield and sensitivity.

4.6 Summary

This chapter centers around the key aspect of the research, which is the introduction and analysis of a novel probing detection concept designed specifically for regular bus lines. The core concept is presented and elaborated upon, encompassing models for both coupled and non-coupled bus lines.

A comprehensive statistical analysis, accompanied by tailored figures of merit, is conducted to assess the effectiveness of the proposed concept. Various parameters and scenarios are taken into consideration to thoroughly examine the efficacy and performance of the concept.

By focusing on the core objective and providing a detailed examination, this chapter lays the foundation for further advancements in probing detection techniques. The analysis presented here serves as a basis for subsequent chapters, which will explore additional aspects and extend the evaluation to encompass irregular bus lines. Through these endeavors, a comprehensive understanding of the proposed concept's capabilities and limitations will be achieved, further contributing to the field of bus line probing detection.

Chapter 5

Different Length Bus Line Wires

5.1 Introduction

In previous analyses of ROPAD [39], the data bus lines were assumed to be regular, conforming to an ideal case [1]. However, in practical on-chip implementations, data bus irregularities can arise due to variations in length and shape, leading to unequal capacitance values in realistic layouts. These irregularities can occur as a result of wire jogging during physical design.

Wire jogging is a technique employed in System-on-Chip (SoC) design to rectify metal routing antenna design rule violations. It involves the placement of physically connected vias between metal layers in the bus layout. Wire jogging can also be used to mitigate crosstalk noise, both within the same metal layer or between different metal layers, by incorporating physical vias. Furthermore, irregularities may arise due to space constraints in the routing process.

Figure 5.1 illustrates an example of wire jogging in a layout, showcasing the proper alignment of cell pins and physically placed vias with the metal lines. As a result, the bus layout exhibits irregularities characterized by discontinuities in the metal line arrangement.

[1] This chapter builds upon the author's publication [40].

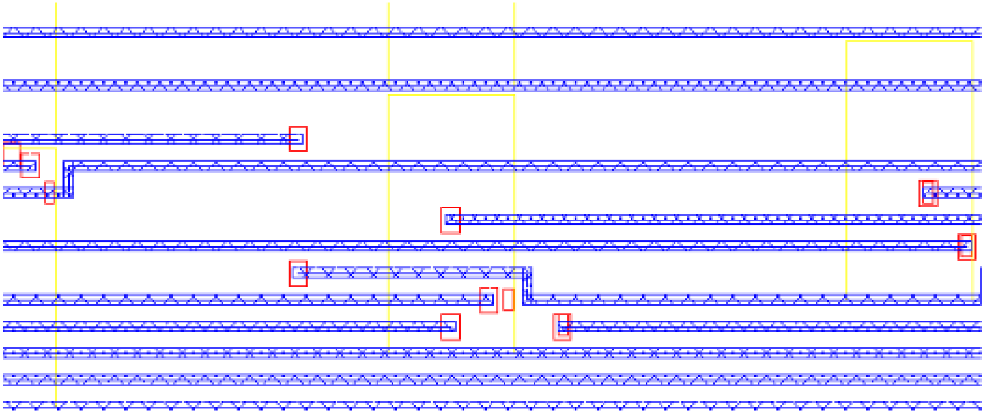


Figure 5.1: Metal route wire jogging in a bus layout.

The irregularity introduced by wire jogging and other layout constraints leads to variations in the length and shape of the bus lines, consequently affecting the capacitance values of each line. As a result, the expected delay differs for each pair of bus lines. To account for this effect, individual values for DB (Delay Bits) can be selected. In the case of an irregular bus with N lines, N DB values are stored in memory.

However, it is important to note that different delays in the bus lines also impact the specific effects of probing. The variations in bus line lengths introduce discrepancies in the counter values employed by ROPAD to detect probing. In order to address this effect, we will now analyze how the counter values used by ROPAD are influenced by the differing lengths of the bus lines.

5.2 Influence on Detection Capabilities

For analyzing the ROPAD behavior in presence of irregularity, we suppose that the first line in the pair has a capacitance value equal to CAP_L and second line has a different capacitance value named as CAP_{L+IR}:

$$CAP_{L+IR} = CAP_L + CAP_{IR}, \tag{5.1}$$

The difference between capacitance values of two lines, corresponds to the irregularity caused capacitance value named CAP_{IR}.

Experimental Investigation

We have analyzed practical usecases in order to define typical values for CAP_{IR} in a normal bus. The lower and upper limits, using ASIC design layout flow for digital logic, was found to be within the range $[-16\%, +16\%]$ of the mean value of the bus line capacitance. In our analysis we limit ourselves to such realistic imbalances of the line capacitances.

Recall that a probe is detected by ROPAD, if attaching a probe causes a sufficient offset in the counter differences. According to alarm generation function, an alarm is triggered if a specific difference of the counter values is reached. As a consequence, reliable probe detection is only possible for counter differences $\Delta C_{unprobed} = C_{u,1} - C_{u,2}$ with no probe attached and counter differences $\Delta C_{probed} = C_{p,1} - C_{u,2}$ with a probe attached, if

$$\Delta C_{probed} - \Delta C_{unprobed}$$

is sufficiently large. Please note, that without loss of generality, we assume the probe to be attached at bus line 1.

Figure 5.2 provides a sweep over $\Delta C_{probed} - \Delta C_{unprobed}$ for different line capacitances of 336 fF, 400 fF and 464 fF and realistic offsets of ±16%. The choice is motivated by the observation from previous research, that probes on larger capacitances are harder to detect and by the maximum targeted line capacitance of 400 fF in our work. The latter results in minimum of 336 fF and a maximum of 464 fF when considering the 16% offset. Please note that for the experiment in Figure 5.2 - in order to decrease the simulation run time - we have taken a smaller reference value for the counters compared e.g. to upcoming experiment in Figure 5.5. That is why the range of $\Delta C_{probed} - \Delta C_{unprobed}$ is different between the two experiments.

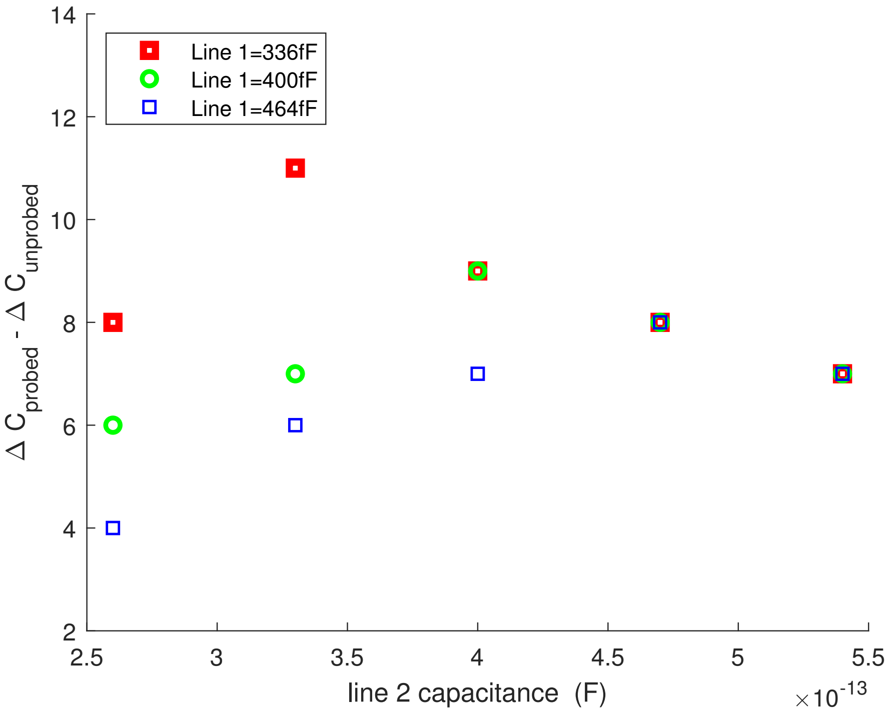


Figure 5.2: Line 1 vs line 2 sweep

In the results, still line 1 is assumed the probed line, and line 2 is the unprobed one. The worst case, i.e., the smallest change in counter values, is observed, if the unprobed line is 16% smaller than the probed one. The case of highest sensitivity against a probe, is reached at the point, where line 1 and line 2 are of equal length, emphasizing the need of considering imbalance in the wire lengths. Please note, that the offset of the counter difference in Figure 5.2 is not symmetric around the maximum. This is because only line 1 is probed, shifting the expected capacitance of line 1 by 20 fF in the probed case.

The general trend allows to conclude, that the worst case for probe detection is

where the bus line has a large capacitance but the unprobed line is shorter than the probed one. A theoretical substantiation of this claim is provided in upcoming Section 5.2. However, the effect also could be explained by the system itself: The cross-coupled counters in ROPAD count until a reference value is reached. A small capacitance has the lowest impact on the counter values if the bus lines are large, and thus the oscillation frequency of the oscillator is low, causing that larger bus lines are harder to protect. However, since the counters stop when the first counter reaches the reference, the faster counter determines the measurement time. The analysis in Figure 4.5 already reveals, that for this reason yield and sensitivity increase when counting for longer.
A reduced capacitance of only one bus line therefore results in a lower measurement time and, consequently, in a less precise resolution of the other bus line's capacitance. A probe attached to the larger bus line is therefore harder to detect in an irregular bus scenario. To take this into account, we therefore attach the probe to the bus line with larger capacitance in the following experiments.

Theoretical Justification

The goal is to substantiate the claim that the best predictability is given for equal length bus lines and that the worst case for detection is a probe attached to the line with higher capacitance. This is another view onto the intuitive description of Figure 5.2 in Section 5.2.

To instantiate the claim, we define function G:

$$\begin{aligned} G &:= \Delta C(C1, C2, C_{REF})_{probed\ state} - \Delta C(C1, C2, C_{REF})_{unprobed\ state} \\ &= \Delta C(C_{L+p}, C_{L+IR}, C_{REF}) - \Delta C(C_L, C_{L+IR}, C_{REF}). \end{aligned}$$

where due to introduction of non-equal length lines, new specific terms are defined to represent the counter values: C_L is the counter value for line 1. The term C_{REF} determines the fact the counters are stopped at this reference value, in each of the probed or unprobed experiments. W.l.o.g, we assume that only line 1 is probed, and that the counter value for line 1 in presence of a probe is C_{L+p}. The irregularity is accounted to the second line which corresponds to counter value $C_{L+IR} = C_L + C_{IR}$ where C_{IR} is the offset.

Note that we use the same notation for the capacitance values: CAP_{L+p} is the capacitance for line 1 in presence of a probe, CAP_L is the capacitance value for line 1, and CAP_{L+IR} is the capacitance value for the second line (the irregular one).

Depending on the values of C_{L+p}, C_L, C_{L+IR} three cases are defined.

Case 1

$CAP_{L+p} > CAP_L > CAP_{L+IR}$

This, represent a part of the curve placed in the left side of its extrema i.e. its maximum point. Replacing the counter values and their differences by Equation (4.10) and Equation (4.7) results in:

$$G = \Delta C(C1, C2, C_{REF})_{probed\ state} - \Delta C(C1, C2, C_{REF})_{unprobed\ state}$$

$$= C_{REF} \cdot (\frac{f_{L+IR} - f_{L+p}}{f_{L+IR}} - \frac{f_{L+IR} - f_L}{f_{L+IR}}).$$

f_{L+p} is the frequency of the ring oscillator corresponding to the C_{L+p}, and in general f_* is the frequency of the ring oscillator corresponding to C_*. C_{REF} is the reference value at which both counters stop.

$$G = C_{REF} \cdot (-\frac{f_{L+p} - f_L}{f_{L+IR}}).$$

According to Equation (4.6), frequencies can be expressed by the capacitance values. We abbreviate the line 1 capacitance without probe CAP_L, the probe capacitance

CAP_P, and the capacitance offset due to irregularity as CAP_{IR}. We suppose that the φ is not changing between the lines, i.e. neglecting possible mismatch. Hence, If φ cancels out, this results in

$$G = \frac{\frac{1}{CAP_L} - \frac{1}{CAP_L+CAP_P}}{\frac{1}{CAP_L+CAP_{IR}}} \cdot C_{REF}$$

$$= C_{REF} \cdot \frac{(CAP_L + CAP_{IR}) \cdot (CAP_P)}{CAP_L \cdot (CAP_L + CAP_P)} \tag{5.2}$$

Taking CAP_{IR} as a parameter, and line 2 capacitance $CAP_L + CAP_{IR}$ as the independent variable in Figure 5.2, the slope is positive for $CAP_L > 0, CAP_P > 0$.

$$\frac{\partial G}{\partial(CAP_L + CAP_{IR})} = C_{REF} \cdot \frac{(CAP_P)}{CAP_L \cdot (CAP_L + CAP_P)} \tag{5.3}$$

I.e. the worst case is reached for the most negative $CAP_L + CAP_{IR}$ in the far left side. The worst case occurs where the G is the minimum in the range, because the possibility of an interference between the distribution increases to its maximum over this point.
The best case is reached if CAP_{IR} approaches 0 (from the left) and C_{L+IR} goes towards C_L. This is the best case in the range, because the possibility of an interference between the distribution decreases to its minimum over this point.

Case 2

$CAP_{L+p} > CAP_{L+IR} > CAP_L$

This, represent a small part of the curve placed in the right side of its extrema i.e. its maximum point where $CAP_{L+p} > CAP_{L+IR}$ is valid.

For this case, the stop criterion changes. In detail: If the wire is not probed, the oscillator causing C_L reaches C_{REF} first and stops the system; If the wire is probed, the oscillator formed by bus line 2 with C_{L+IR} reaches C_{REF} first and stops the system. This is reflected in Equation (4.7) by replacing the divider for the two different cases by the frequency determining the stop criterion:

$$G = \Delta C(C1, C2, C_{REF})_{probed\ state} - \Delta C(C1, C2, C_{REF})_{unprobed\ state}$$

$$= C_{REF} \cdot (\frac{f_L - f_{L+IR}}{f_L} - \frac{f_{L+p} - f_{L+IR}}{f_{L+IR}})$$

$$= C_{REF} \cdot (\frac{-f_L \cdot f_{L+p} - f_{L+IR} \cdot f_{L+IR}}{f_L \cdot f_{L+IR}} + 2)$$

$$= (\frac{-\frac{1}{CAP_L+CAP_P}}{\frac{1}{CAP_L+CAP_{IR}}} - \frac{\frac{1}{CAP_L+CAP_{IR}}}{\frac{1}{CAP_L}} + 2) \cdot C_{REF}$$

$$= (-\frac{CAP_L + CAP_{IR}}{CAP_L + CAP_P} - \frac{CAP_L}{CAP_L + CAP_{IR}} + 2) \cdot C_{REF}$$

$$\frac{\partial G}{\partial (CAP_L + CAP_{IR})} = (-\frac{1}{CAP_L + CAP_P} + \frac{CAP_L}{(CAP_L + CAP_{IR})^2}) \cdot C_{REF} \qquad (5.4)$$

Taking an approximation $\lim_{CAP_{IR} \to 0}(CAP_L + CAP_{IR}) = CAP_L$ for simplification:

$$\frac{\partial G}{\partial (CAP_L + CAP_{IR})} = (-\frac{1}{CAP_L + CAP_P} + \frac{1}{(CAP_L + CAP_{IR})}) \cdot C_{REF} \qquad (5.5)$$

The slope with respect to $CAP_L + CAP_{IR}$ is negative for $CAP_{IR} < CAP_P, CAP_L > 0, CAP_P > 0$, i.e., in the complete defined region for case 2, in the right side of the maximum point in the curve. In the discussed range, by increasing the line 2 capacitance $CAP_{L+IR} = CAP_L + CAP_{IR}$, the term G decreases due to the negative slope of the curve. Note, that for $CAP_{IR} \to 0$ the result is the same as for Eq. 5.2 above and for $CAP_{IR} \to CAP_P$ the result is the same as for Eq. 5.6 below proofing continuity.

Case 3

$CAP_{L+IR} > CAP_{L+p} > CAP_L$

This, represent the remaining part of the curve placed in the far right side of its extrema i.e. its maximum point where $CAP_{L+IR} > CAP_{L+p}$ is valid.
Equivalent to the previous findings,

$$G = \Delta C(C1, C2, C_{REF})_{probed\ state} - \Delta C(C1, C2, C_{REF})_{unprobed\ state}$$

$$= C_{REF} \cdot \left(\frac{f_L - f_{L+IR}}{f_L} - \frac{f_{L+p} - f_{L+IR}}{f_{L+p}} \right)$$

$$= C_{REF} \cdot \left(\frac{f_{L+IR} \cdot f_L - f_{L+IR} \cdot f_{L+p}}{f_L \cdot f_{L+p}} \right)$$

$$= C_{REF} \cdot \left(\frac{f_{L+IR}}{f_{L+p}} - \frac{f_{L+IR}}{f_L} \right)$$

$$= C_{REF} \cdot \left(\frac{CAP_L + CAP_P}{CAP_L + CAP_{IR}} - \frac{CAP_L}{CAP_L + CAP_{IR}} \right)$$

$$= C_{REF} \cdot (\frac{CAP_P}{CAP_L + CAP_{IR}}) \tag{5.6}$$

$$\frac{\partial G}{\partial (CAP_L + CAP_{IR})} = (-\frac{CAP_P}{(CAP_L + CAP_{IR})^2}) \cdot C_{REF} \tag{5.7}$$

The slope with respect to line 2 capacitance $CAP_L + CAP_{IR}$ is negative for $CAP_L > 0$ and $CAP_P > 0$, i.e., for the defined region. The maximum of G in the range, is hence reached at the border transition point between the case 2 and case 3. At this point, CAP_{IR} is at its smallest value in the defined range. Further, the worst case in this region (smallest counter difference) is reached for the largest value of $CAP_L + CAP_{IR}$.

Where the worst case and best case have been declared previously within Case 1. Let A be the absolute value of CAP_{IR}. Comparing the results of case 1 and case 3 with $CAP_{IR} = A$ in Equation (5.6) and $CAP_{IR} = -A$ in Equation (5.2) results (after removal of C_{REF}) in

$$(\frac{CAP_P}{CAP_L + A}) > \frac{(CAP_L - A) \cdot (CAP_P)}{CAP_L \cdot (CAP_L + CAP_P)}$$

since

$$1 > \frac{CAP_L^2 - A^2}{CAP_L \cdot (CAP_L + CAP_P)}$$

holds for our case, namely $CAP_L > A > 0$ and $CAP_P > 0$. This proves that for same absolute irregularity, probing on the slower line (case 1) results in a worse case.

5.3 Limitation of the Previous Approach

Increasing the reference counter in the system might be seen as a straight forward solution to counter the difficulty of observing probes on unbalanced bus lines. If this would work, a single distinguisher (DB) would suffice. However, realistic simulation results show that such a simple approach is impractical.

Indeed, with unpredictable local variations considered, setting a "global DB" according to Equation (4.13) results for large imbalances in low reliability and yield. To show this effect, we performed Monte Carlo analyses regarding local process variations under worst case conditions over several reference counter values C_{REF}.
Sensitivity and reliability are calculated under a fixed global DB.
We showcase two experiments:
(i) Probing of an irregular bus of nominal capacitance of 100 fF±16%, i.e., the faster (unprobed) bus line is set to 84 fF, whereas the bus line under attack is set to nominal capacitance of 116 fF (before probing). (ii) Probing is also considered for a case of a bus line pair targeting 400 fF±16%. I.e., the smaller bus line is modeled to have 336 fF; the large bus line is under attack and is modeled to have a capacitance of

464 fF. Results for the two cases are shown in Figure 5.3 and, respectively, Figure 5.4. For the shorter irregular bus line defined in case (i), it can be seen in Figure 5.3 that indeed detection of the probe is possible and both sensitivity and yield limiting close to 100%, just as it was the case for bus lines with equal length shown previously in Figure 4.5.

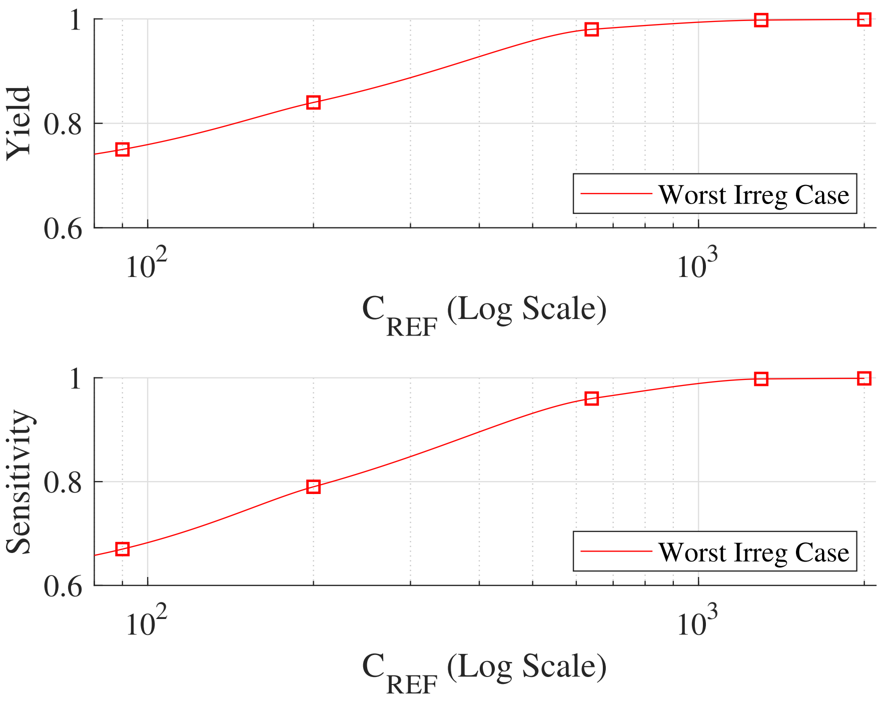


Figure 5.3: Sensitivity and Yield for the worst case of 116 fF vs. 84 fF using the global *DB* according to Equation (4.13)

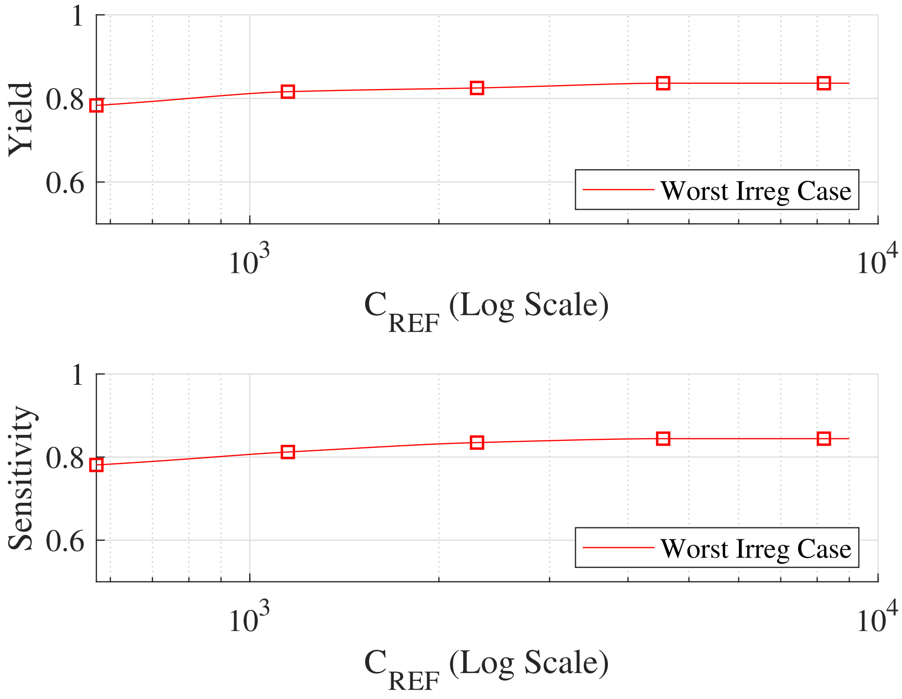


Figure 5.4: Sensitivity and Yield for the worst case of 464 fF vs. 336 fF using the global DB according to Equation (4.13)

However, in case (ii), for longer irregular bus lines the approach of increasing the reference counter does no longer show significant effect. This is visible in Figure 5.4. In this case, yield and sensitivity remain below 90% even for large values of the reference counter. Due to these practically too low values, post manufacturing calibration is suggested in this work in order to decide if a bus line is probed.

5.4 Calibration

In order to make it possible to use a single value global DB, we need to improve the sensitivity and yield for the global DB. For this purpose we first need to analyze how much the process variation could reduce the figures. Afterwards, we suggest a new decision criterion to distinguish probed from unprobed cases.

Impact of Process Variations on Probe Detection

As previously discussed, with a single decision bound yield and sensitivity are too low for realistic measurement times (counter references) and large bus lines. The reason for this is that due to the imbalance the distance $\Delta C_{probed} - \Delta C_{unprobed}$ becomes smaller.
As a consequence, the distributions for probed and unprobed bus lines (named Y and Z in Figure 4.4) get closer. For a fixed DB placed according to Equation (4.13), yield and sensitivity as defined in Equation (4.15) and Equation (4.17) are reduced.

To find a distinguisher, which allows a decision if finding a probe is still possible, we have measured the distribution of the difference of counter differences $\Delta C_{probed} - \Delta C_{unprobed}$ between probed and unprobed samples.
The result from Monte Carlo analyses with respect to local process variations is exemplified in Figure 5.5 for the case of a 464 fF probed line and a 336 fF unprobed line.

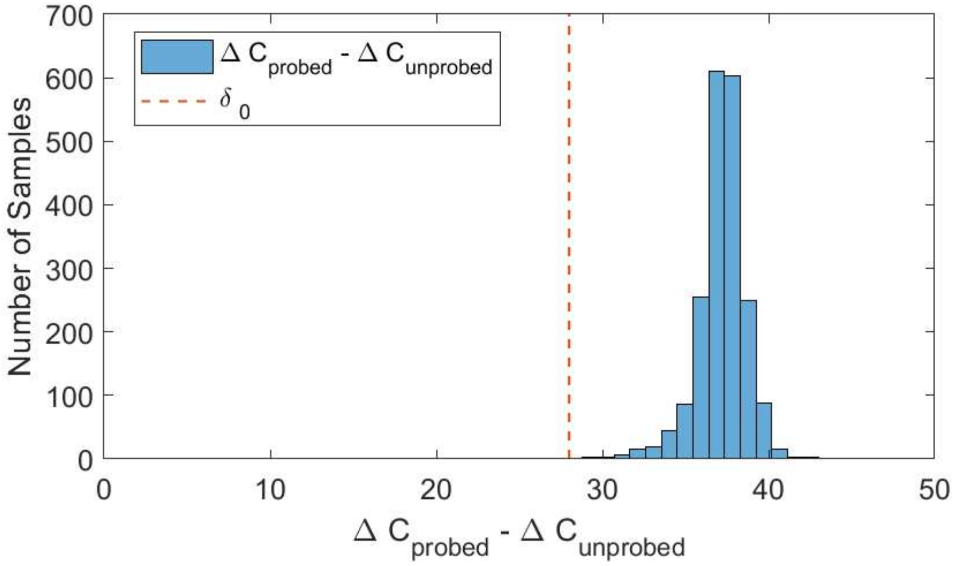


Figure 5.5: Distribution of difference between 20fF probed and unprobed samples at the worst design corner. This refers to a case with $L1 = 464fF$ and $L2 = 336fF$ and $C_{REF} = 4894$, Data is generated out of 2000 Monte Carlo samples.

For each Monte Carlo sample, i.e., for each simulated instance of the circuit, $\Delta C_{probed} - \Delta C_{unprobed}$ was computed. Obviously, the counter differences are distributed according to some near-Gaussian distribution. Distinguishing probed and unprobed state is theoretically possible as long as $\Delta C_{probed} - \Delta C_{unprobed} > 0$.[2]

The measurements for this case have been done for a counter reference of 4894 resulting in a yield of 0.82 and a sensitivity of 0.83 for using a single DB according to Equation (4.13). Obviously, the situation can be improved, when the decision is related to the *expected* result of ΔC_{probed}:
After manufacturing, local variations causing the detection problems are fixed for the physical metal line of a wire as well as for the bus buffers. An additionally attached probe shifts the counter output for similarly sized bus lines by a similar amount. Therefore, given the expectation of a counter without probe and the ex-

[2] Strictly spoken $\Delta C_{probed} - \Delta C_{unprobed} \neq 0$ would suffice, but since the goal is a one-sided test, we restrict ourselves to > 0.

pected deviation for this class of similar bus lines when probed suffices for detection.

5.5 A New Decision Boundary

The problem that different length bus lines are harder to distinguish causes that more precise knowledge about the line delay differences is needed. A straight forward solution to overcome this problem is to measure the line delay caused counter differences post manufacturing in an unprobed setting and with an emulated probe attached.

While this approach allows to setup the best possible distinguisher (DB) for each line pair, it also comes with practical downsides. The emulation of the probe requires the implementation of a relatively large capacitance.

First, this causes area overhead and also increases calibration time since measurement with and without emulated probe attached are needed. Second, this capacitance must be designed, placed, and routed so that for all line pairs, which must be calibrated, the seen emulated probe capacitance reflects very precise the value of the targeted probe, e.g., 20 fF. In particular, if lines have a too large emulated probe attached during calibration, the detection capability of actual probes is reduced. In addition an attacker might try to misuse the capacitance in order to get advantage for an attack.

Therefore we propose a method without emulation of the probe during calibration. I.e., we only involve self-calibrating for the unprobed wires and find limits for the probed case based on our pre-silicon Monte Carlo results.

It was already obvious from Figure 5.5 that realization of process variations influences the unprobed wire and the probed one similarly. In particular, for all considered wire lengths and for all process variations sampled, the difference of counter differences with and without probe attached is larger than a specific offset value named as δ_0:

$$\Delta C_{probed} - \Delta C_{unprobed} > \delta_0 \tag{5.8}$$

This δ_0 defines the difference between probed and unprobed Line 1. The impact of the choice of this value is as follows: Assuming no noise or aging and onboard calibration under worst case conditions, an arbitrarily small $1 \geq \delta_0 > 0$ would result in 100% yield – since all other effects are covered by the worst case assumption – and extremely high sensitivity. However, tiniest variations of the oscillator would cause yield loss. Shifting δ_0 towards larger values decreases the sensitivity.

Let from now on δ_0 be selected so that it is the largest possible value, for which the required sensitivity (e.g., $sens > 99.9\%$) is met. We suggest that this value of δ_0 is computed during design time and permanently stored to the device in a way that cannot be manipulated by the attacker.[3]

With $\Delta C_{cal,un}$ the actual realization of on-chip calibrated counter difference values for two unprobed bus lines, $\Delta C_{cal,un} + \delta_0$ defines the value where latest an alarm must be triggered to fulfill sensitivity requirements.

However, in real scenarios this value can shift, e.g., due to noise or aging. Therefore, we define the new decision boundary $CALDB$ to decide if an observed counter difference value corresponds to a bus under attack, as

$$CALDB = \Delta C_{cal,un} + \frac{\delta_0}{2} \tag{5.9}$$

$\frac{\delta_0}{2}$ accounts for placing the $CALDB$ in between the probed and unprobed distributions, to achieve the best noise margin. $CALDB$ and its positioning halfway between the unprobed distribution and the δ_0 is visualized in Figure 5.6. Please note that in this experiment - compared to the one in Figure 4.4 - different conditions are taken into account, e.g. in this case in Figure 5.6, different counter reference value, and different line capacitance is taken into account.

[3] Please note, that the value does not carry secret information, i.e., it might be readable.

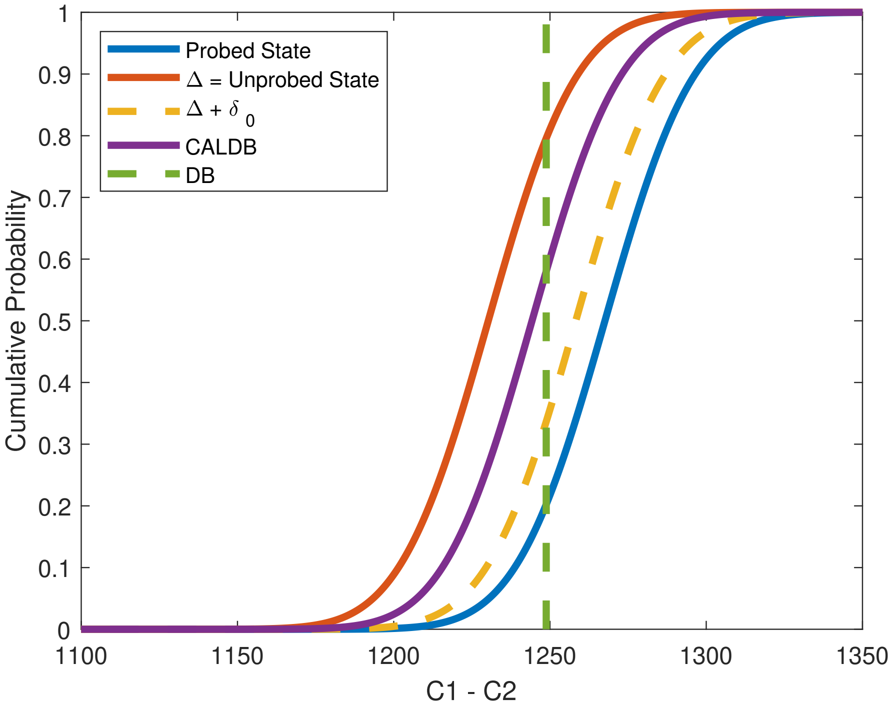


Figure 5.6: δ_0, CALDB, DB and the samples distribution of probed and unprobed cases.This refers to a case with $L1 = 464fF$ and $L2 = 336fF$ and $C_{REF} = 4894$, Data is generated out of 2000 Monte Carlo samples.

The halfway positioning ensures the best possible sweet spot for achieving the highest yield and sensitivity.

The figure also shows the benefit of $CALDB$ over the previously used DB: While the previous DB (green vertical line) stands for a fixed value, so that counter differences which are already initially close to the boundary are less reliable, $CALDB$ follows this distribution of the counter difference for the unprobed state (red line), so that the probed state can be distinguished from the unprobed state even for extreme local process variations. The previous DB (green vertical line) provides only sensitivity and yield of 80% for this worst irregular case, however the new $CALDB$ provides much higher values, i.e. both are $\geq$ 99.99%.

5.6 Secondary Effects

After introducing the new $CALDB$ as a new distinguisher, we now discuss several properties of the new version of ROPAD.

Measurement Time

The discussion until here raises the question, how small the measurement time can get, when using $CALDB$. Since measurement time in ROPAD is determined by C_{REF}, we designed an experiment - with the worst case of 464fF vs. 336fF - to show how δ_0 increases over C_{REF} increase. Results are shown in Figure 5.7.

Increase of C_{REF} means longer counting of the counters. This corresponds to a longer test duration but also to a better resolution of possible differences between bus lines. $\delta_0 > 0$ is a required condition to achieve the maximum yield which is achievable for each $C_{REF} > 1000$ equal to $testtime > 600ns$. By higher δ_0 we achieve a better separation of the probed and unprobed samples, and therefore gain more space for placing the $CALDB$. This is equivalent to increasing sensitivity and yield.

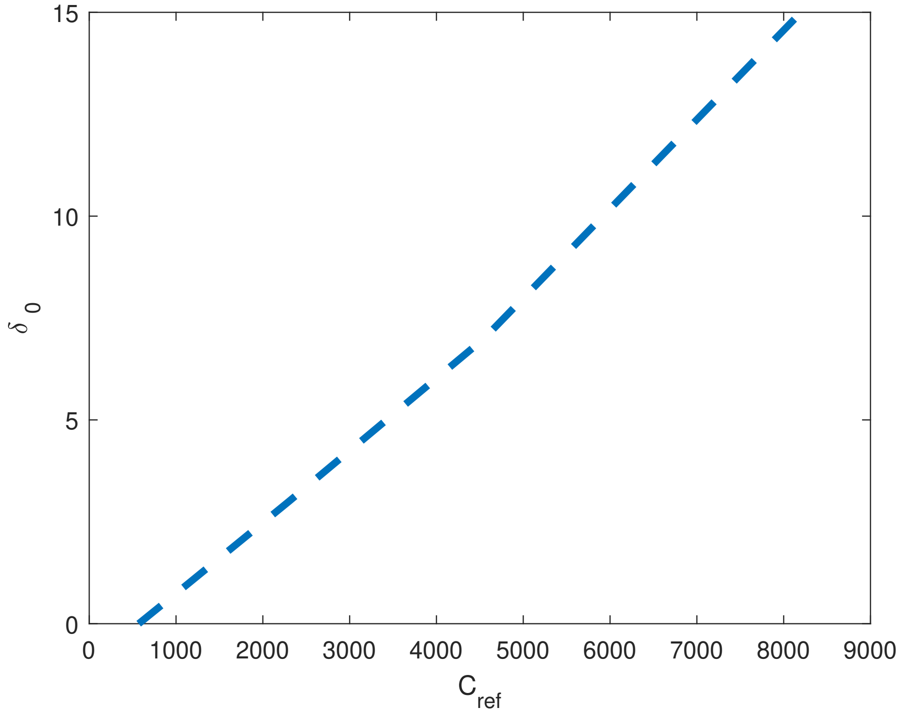


Figure 5.7: δ_0 increases over C_{REF} increment, data refer to the worst case of 464fF vs. 336fF.

Influence of Jitter

Global effects like temperature effects cancel out to a large extend due to the differential measurement in ROPAD and are already considered by the worst-case assumptions above.

Effects like thermal noise causing jitter, however, can influence the frequency of each oscillator individually. Therefore, we discuss the influence of jitter on ROPAD's performance as a last remark in this work.

Figure 5.8 exemplifies how jitter might influence the counter value difference $\Delta C_{cal,un}$. The experiment is performed with the same conditions defined in Figure 5.6. Absolute jitter is additionally simulated here, using the Cadence jitter simulation software [4]. Based on the results we see that 8 cycle shifts is a realistic scenario.

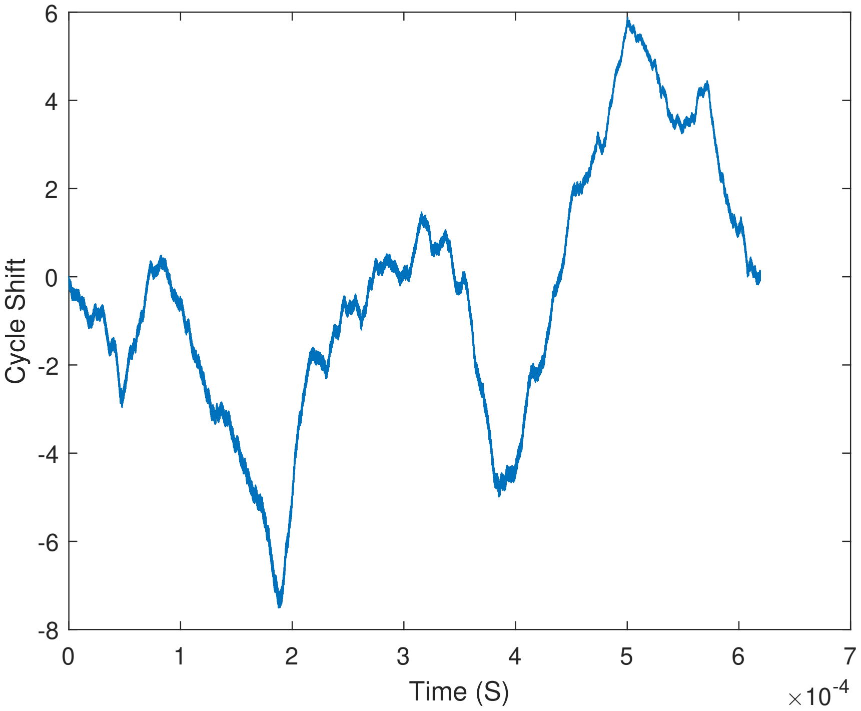


Figure 5.8: cycle shift occurring in the counter difference, due to absolute jitter. The X axis is the transient simulation time. A random jitter noise has been inserted by the simulation tool on top of the transient simulation.

Although jitter might average out over long measurement periods, it is observable that it might still affect the least significant bits of the counter difference. The precise amount of jitter is defined via the variance of the noise and cannot be generalized across technologies and designs. Nevertheless, inverter gates optimized for noise reduction, like they are used in our design, help to limit the absolute jitter.

To consider jitter in our design, it suffices to predict a value δ_J that is with sufficient probability larger than the effect of jitter on the probed and the unprobed

counter value, e.g. in this experiment a value larger than 8 must be chosen.
With an appropriate choice of δ_0, in particular

$$\delta_0 > 2\delta_J$$

also this noise effect is taken into account. Therefore, in this experiment $\delta_0 > 16$ must be considered.

The remaining effect, we leave for future work is aging. That effect is irreversible and would not cancel out. While still a sufficiently long measurement period and, therefore, a large value of δ_0 might result in acceptable yield and sensitivity, another option is available with the new version of ROPAD:
If the device can be brought into a trusted environment for maintenance, recalibration is possible.

5.7 Comparison to the Previous Approach

Table 5.1 presents a comprehensive comparison of the state-of-the-art digital PADs. In contrast to CaLIAD, which is specifically designed for regular buses, ROPAD offers protection for both regular and irregular buses within the specified range. ROPAD surpasses CaLIAD in terms of its ability to detect probes, accommodating regular and irregular bus lines with a nominal capacitance of up to 400 fF. Notably, this capacitance threshold is four times higher than that considered by CaLIAD.

Moreover, as discussed in [39], ROPAD exhibits a smaller footprint in terms of area utilization. The compact design of ROPAD enables efficient integration within the chip layout, optimizing the utilization of available resources. This characteristic further highlights the superiority of ROPAD as a versatile and space-efficient solution for bus protection.

Table 5.1: FoM comparison for SoA of digital PADs

	ROPAD	CaLIAD[51]
Irregular bus Protection	Yes	No
Max. CAP_{IR}[fF] [$Irreg.$]	$(-64\text{ fF} + 64\text{ fF})$	n/a
Max. CAP_L [fF]	400	100
Area[GE]	181	352
Protected Memory [DB](Reg.)	1	N ($2N$)
Protected Memory [DB](Irreg.)	N	n/a
in-Si Cal. Runs [$Reg.$]	0 (1)	$N(2N@CAP_P)$
in-Si. Cal. Runs [$Irreg.$]	N (0)	n/a

The memory requirements for ROPAD in terms of protected memory are notably lower compared to CaLIAD. For regular bus lines, ROPAD only necessitates the storage of a single decision bound. Conversely, CaLIAD imposes a more demanding storage requirement, as a minimum of N different calibration values must be stored. If CaLIAD is calibrated under an emulated probe to achieve optimal sensitivity and yield, this number increases to $2N$.

One significant advantage of ROPAD over CaLIAD is its ability to address irregular bus lines by utilizing N stored distinguishers. This capability sets ROPAD apart, as CaLIAD does not account for such cases in its design.

Furthermore, ROPAD outperforms CaLIAD in terms of in-silicon calibration runs. In the case of regular bus lines, ROPAD does not require calibration, although incorporating a calibration run can enhance reliability. As for irregular bus lines, ROPAD employs calibration based on N line pairs exclusively in the unprobed state, without the need for an emulated probe capacitance. However, for shorter line lengths, calibration is unnecessary for irregular bus lines as well.

It is worth noting that ROPAD achieves exceptional sensitivity and yield, nearing 100%, for both regular and irregular bus lines. This remarkable performance underscores the effectiveness of ROPAD as a reliable and high-performing solution for bus protection.

5.8 Summary

In this chapter, we have expanded upon the capabilities of ROPAD by enhancing its ability to detect probes on irregular data buses. This represents a significant advancement over its previous version, which was primarily focused on detecting probes on regular buses [39].

The working conditions of the detector IP have been thoroughly analyzed, taking into account both nominal and detailed statistical considerations. In order to achieve optimal yield and sensitivity, ROPAD has been configured with a post-silicon calibration approach, prioritizing minimal design overhead and latency.

To this end, we have proposed a statistically driven decision boundary and have explored its application in scenarios involving irregularity, local variation, and jitter. Through extensive evaluation, we have demonstrated that the new approach is well-suited for detecting probes on irregular bus lines, with an acceptable overhead. Furthermore, the enhanced ROPAD remains capable of detecting attackers' probes even at small capacitance values as low as 20 fF.

These findings highlight the effectiveness and practicality of the proposed methodology, showcasing its potential for reliable and accurate detection of probe attacks on irregular data buses. The improved capabilities of ROPAD open up new possibilities for robust security measures in modern electronic systems.

Chapter 6

Summary and Outlook

Throughout this work, the primary focus has been on addressing invasive attacks and developing effective countermeasures through the design of a detection circuitry, namely "A Microprobing Detector".

The limitations of traditional time-to-digital converter-based self-tests, which serve as the state-of-the-art probing detectors, have been thoroughly discussed in this thesis. To address these limitations, a different measurement principle with reduced overhead and calibration requirements has been employed. Novel probing detection mechanisms have been proposed, along with a low-effort calibration scheme aimed at increasing the noise margin and reducing test time.

For regular bus lines, the protected bus capacitance has been increased by a factor of four compared to existing approaches. Furthermore, for the first time, it has been demonstrated that the proposed circuitry is capable of effectively protecting non-equal bus line lengths, referred to as irregular lines. This is of utmost importance as non-equal length bus lines are prevalent in the industry, driven by the need to mitigate antenna effects, cross-talk, and employ wire jogging and inter-metal layer routing techniques.

In addition, the effects of coupling on bus lines, which can diminish the sensitivity and yield of the detection core, have been analyzed for the first time. Coupling is a significant real-life issue in the industry, and understanding and addressing its impact is crucial for ensuring acceptable industrial production standards. The proposed approach has been shown to exhibit resilience against coupling effects to an acceptable level.

Furthermore, a noise margin has been introduced, and the probing detection core has been analyzed with respect to jitter noise, a factor that can impact its performance. These considerations contribute to the overall robustness and reliability of the detection system.

As semiconductor technology continues to advance with shrinking process sizes and increasing clock frequencies, the impact of aging on the design and sign-off of application-specific integrated circuits becomes increasingly significant. Aging poses a considerable challenge in the design and manufacturing of ASICs, as it can lead to performance degradation, reliability issues, and increased power consumption over time. To address this concern, it is essential to implement robust mechanisms that can accurately identify and isolate the effects of aging from other sources of variation. The development of such strategies requires a comprehensive understanding of the underlying aging mechanisms and their distinct characteristics. By leveraging this knowledge, designers can devise detection techniques that are specifically tailored to differentiate aging effects from other variations, such as process variations or environmental factors. Moreover, these strategies must be implemented with careful consideration of the die size. Given the limited physical space available on a semiconductor die, it is crucial to minimize the footprint of the aging detection mechanisms. This entails efficient utilization of resources, intelligent circuit design, and optimization of detection algorithms to ensure the least possible impact on the overall die size. However, traditional Vernier time-to-digital converter-based built-in self-tests suffer from notable calibration and post-processing overhead, limiting their effectiveness in addressing aging concerns. These built-in self-tests also impose challenges in terms of extensive metal routing and the incorporation of large multiplexer (MUX) switches, which introduce readout errors to the measurement core. Furthermore, the aging stress on the primary time-to-digital converter sensors needs to be carefully managed, particularly when dealing with an expanding number of critical paths.

To overcome these limitations, a novel approach has been proposed to reduce overhead by leveraging the frequency dependence of the negative bias temperature instability aging mechanism. By exploiting the frequency dependency of aging, significant gains in terms of area and power consumption, exceeding 90%, can be achieved. This reduction in overhead not only addresses the calibration and post-processing challenges associated with traditional BISTs but also alleviates the aging stress on crucial components such as the time-to-digital converter sensors. Consequently, the proposed approach enables designers to enhance the overall reliability and performance of ASICs in the presence of aging effects.

The utilization of the frequency dependence of the NBTI aging mechanism represents a promising solution for managing the challenges posed by aging in modern ASIC designs. By incorporating this understanding into the design process, designers can effectively mitigate the negative impact of aging while optimizing resource utilization. This approach ensures the longevity and reliability of ASICs even as process sizes continue to shrink and transistor count, compute power as well as clock frequencies push technological boundaries.

To safeguard the calibration data of aging sensors from potential cloning attacks, a novel approach utilizing physical unclonable functions has been adopted. PUFs present a more efficient alternative to memory encryption, ensuring the integrity and security of the sensitive calibration data.

It is important to recognize that aging not only poses challenges for application-specific integrated circuits but also affects the performance of physical unclonable functions. Previous investigations have revealed that aging can induce sign bit flips in PUFs, rendering them unreliable and falling out of the acceptable yield range for their intended application. This phenomenon underscores the need to address the issue and enhance the reliability of PUFs against the detrimental effects of aging.

In this regard, a proposal to improve the stability and longevity of ring oscillator-based PUFs in the face of aging has been presented and discussed. The objective is to devise measures that can effectively counteract the impact of aging on PUFs, ensuring their reliable operation throughout their lifespan. By enhancing the resilience of

PUFs against aging effects, the overall reliability and dependability of these security components can be significantly improved.

The proposed approach not only addresses the specific challenges associated with aging in PUFs but also aligns with the broader objective of maintaining the trustworthiness and effectiveness of these critical security elements. By ensuring the stability and reliability of PUFs over their operational lifetimes, the potential risks arising from aging-related issues can be mitigated, bolstering the overall security posture of the system.

In conclusion, the selected approach of utilizing physical unclonable functions as a safeguard for calibration data, or any other type of sensitive information, represents a strategic choice that brings about a more streamlined and efficient solution compared to traditional memory encryption methods. By leveraging the unique characteristics of PUFs, the integrity and security of crucial data can be effectively protected.

Moreover, the proposal to enhance the reliability of ring oscillator-based PUFs against the deteriorating effects of aging holds significant importance in achieving prolonged stability and dependability in the functioning of these essential security components. Aging poses a substantial risk not only to application-specific integrated circuits but also to the reliability and performance of PUFs themselves. Addressing this concern through measures that enhance the resilience of PUFs against aging effects is a crucial step toward ensuring their long-term viability and effectiveness.

The adoption of PUFs as a protective measure demonstrates a forward-thinking and innovative approach in the field of data security. By leveraging the inherent uniqueness of physical characteristics, PUFs offer a robust and efficient defense against various attacks, including cloning attempts. Compared to conventional memory encryption methods, the utilization of PUFs provides a more streamlined and resource-efficient solution, reducing processing overhead while ensuring the integrity and confidentiality of critical data.

Furthermore, the focus on enhancing the reliability of ring oscillator-based PUFs against aging effects underscores the commitment to long-term stability and dependability. Aging-induced sign bit flips in PUFs can jeopardize their performance and render them unreliable for their intended purposes. Thus, the proposed measures aim to mitigate these effects, enabling PUFs to maintain their functionality and effectiveness over an extended operational lifespan.

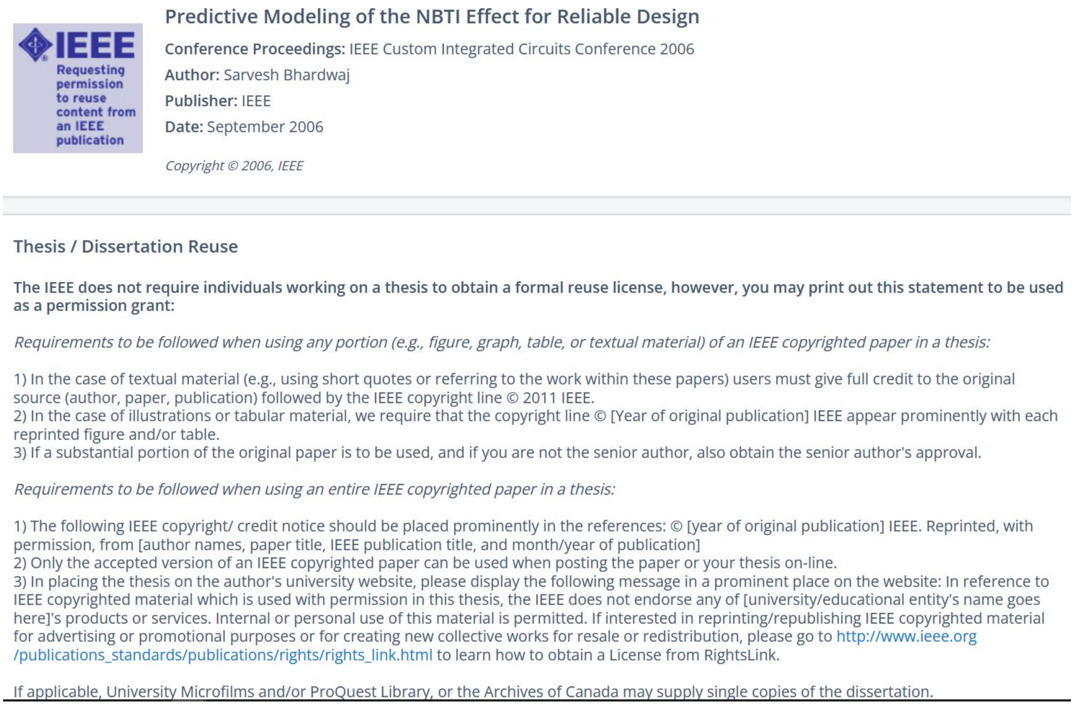
In summary, the deployment of PUFs as a protective measure for calibration data, along with the effort to bolster the reliability of ring oscillator-based PUFs against aging effects, showcases a comprehensive and forward-looking approach to data security. By leveraging innovative techniques and ensuring long-term stability, these critical security components can effectively protect sensitive information while maintaining their resilience and performance against evolving threats.

The proposed approaches presented in this thesis have the potential to greatly facilitate the integration of probing detection cores into industrial-grade application-specific integrated circuits. By addressing key limitations and enhancing the reliability, robustness, and security aspects, these advancements pave the way for more secure and trustworthy electronic systems in various domains. ■

Copyright Notes

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Technical University of Munich's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to: `http://ieee.org/publications_standards/publications/rights/rights_link.html` to learn how to obtain a License from RightsLink.

This is in accordance with the disclaimer from IEEE in fig. 1.


IEEE
Requesting permission to reuse content from an IEEE publication

Predictive Modeling of the NBTI Effect for Reliable Design

Conference Proceedings: IEEE Custom Integrated Circuits Conference 2006
Author: Sarvesh Bhardwaj
Publisher: IEEE
Date: September 2006

Copyright © 2006, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

Figure 1: IEEE copyright note [5]

Bibliography

[1] Accessed: 2023-01-09. URL: https://www.mos-ak.org/munich_2013/presentations/05_Leonhard_Heiss_MOS-AK_Munich_2013.pdf.

[2] Accessed: 2022-11-25. URL: https://embeddedcomputing.com/technology/debug-and-test/emulation-tools-reinvigorate-power-analysis-in-large-socs.

[3] Behzad Razavi, The Bandgap Refernce, A Circuit for all Seasons, IEEE Solid-state Circuits Magazine, Summer 2016, Accessed: 2023-01-18. URL: http://www.seas.ucla.edu/brweb/papers/Journals/BRSummer16Bandgap.pdf.

[4] Accessed: 2022-01-17. URL: https://www.cadence.com/ko%5C_KR/home/tools/custom-ic-analog-rf-design/circuit-simulation/spectre-simulation-platform.html.

[5] Accessed: 2022-08-22. URL: https://s100.copyright.com/AppDispatchServlet.

[6] K. Hofmann et al. »Highly accurate product-level aging monitoring in 40nm CMOS«. In: *2010 Symposium on VLSI Technology, Honolulu, 2010, pp. 27-28.*

[7] Nasim Pour Aryan et al. »In situ measurement of aging-induced performance degradation in digital circuits«. In: *2016 21th IEEE European Test Symposium (ETS)*. 2016, pp. 1–2.

[8] A. Balankutty et al. »Mismatch Characterization of Ring Oscillators«. In: *IEEE Custom Integrated Circuits Conference*. 2007.

[9] Thomas Baumann et al. »On-chip self calibrating delay monitoring circuitry«. Patent US8228106B2. May 2013.

[10] Sarvesh Bhardwaj et al. »Predictive Modeling of the NBTI Effect for Reliable Design«. In: *IEEE Custom Integrated Circuits Conference 2006*. 2006, pp. 189–192. DOI: 10.1109/CICC.2006.320885.

[11] K. A. Bowman. »Energy-efficient and Metastability-Immune Timing-Error Detection and Recovery Circuits for Dynamic Variation Tolerance«. In: *ISSCC*. 2008.

[12] K. A. Bowman et al. »A physical alpha-power law MOSFET model«. In: *International Symposium on Low Power Electronics and Design*. 1999.

[13] N. Callegari. »Path Selection for Monitoring Unexpected Systematic Timing Effects«. In: *Proc. ASP-DAC, 2009*.

[14] M. Chen. »"A TDC-based test platform for dynamic circuit aging characterization,"« in: *2011 International Reliability Physics Symposium, Monterey, CA, 2011, pp. 2B.2.1-2B.2.5*.

[15] A. Drake. »A Distributed Critical-path Timing Monitor for a 65nm High-performance Microprocessor«. In: *Proc. Digest of Technical Papers, IEEE ISSCC 2007*. Feb. 2007.

[16] F. Firouzi. »Representative Cirtical-Path Selection for Aging-Induced Delay Monitoring«. In: *IEEE ITC, 2013*.

[17] D. Ganta and L. Nazhandali. »Study of IC aging on ring oscillator physical unclonable functions,« in: *Fifteenth International Symposium on Quality Electronic Design, Santa Clara, CA, 2014, pp. 461-466*.

[18] Stephan Henzler. »Time-to-Digital Converter Basics«. In: *Time-to-Digital Converters*. Springer, 2010.

[19] A. Herrmann et al. »Bringing Analog Design Tools to Security: Modeling and Optimization of a Low Area Probing Detector«. In: *International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design*. 2018.

[20] M. H. Hsieh. »The impact and implication of BTI/HCI decoupling on ring oscillator«. In: *2015 IEEE International Reliability Physics Symposium, Monterey, CA, 2015, pp. 6A.4.1-6A.4.5.*

[21] Y. C. Huang. » "Delay effects and frequency dependence of NBTI with submicrosecond measurements," « in: *2015 IEEE International Reliability Physics Symposium, Monterey, CA, 2015, pp. 4A.2.1-4A.2.5.*

[22] V. Immler et al. »B-TREPID: Batteryless tamper-resistant envelope with a PUF and integrity detection«. In: *IEEE International Symposium on Hardware Oriented Security and Trust.* 2018.

[23] Yuval Ishai, Amit Sahai and David Wagner. »Private Circuits: Securing Hardware against Probing Attacks«. In: *Advances in Cryptology.* 2003.

[24] D. Persaud J. Keane X. Wang and C. H. Kim. »An All-In-One Silicon Odometer for Separately Monitoring HCI, BTI, and TDDB«. In: *IEEE Journal of Solid-State Circuits.* 2010.

[25] K. Kang. »NBTI Induced Performance Degradation in Logic and Memory Circuits: How Effectively Can We Approach a Reliability Solution?« In: *Proc. Asia/South Pacific Desgin Autom. Conf. (ASP-DAC).* 2008.

[26] E. Karl et al. »Compact In-Situ Sensors for Monitoring Negative-Bias-Temperature-Instability Effect and Oxide Degradation«. In: *2008 IEEE International Solid-State Circuits Conference - Digest of Technical Papers.* 2008, pp. 410–623. DOI: 10.1109/ISSCC.2008.4523231.

[27] T. H. Kim. »"Silicon Odometer: An On-Chip Reliability Monitor for Measuring Frequency Degradation of Digital Circuits," « in: *IEEE Journal of Solid-State Circuits, vol. 43, no. 4, pp. 874-880, April 2008.*

[28] T. T. H. Kim. »A Ring-Oscillator-Based Reliability Monitor for Isolated Measurement of NBTI and PBTI in High-k/Metal Gate Technology«. In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 23, no. 7, pp. 1360-1364, July 2015.*

[29] N. Kimizuka. »Impact of Bias Temperature Instability for Direct Tunneling Ultrathin Gate Oxide on MOSFET Scaling«. In: *Dig. Tech. Papers-Symp. VLSI Technology*. 1999.

[30] J. Fahrny M. T. Rahman D. Forte and M. Tehranipoor. »AROPUF An aging-resistant ring oscillator PUF design,« in: *2014 Design, Automation and Test in Europe Conference and Exhibition.*

[31] S. Lange M. Zorzi A. Gluhak and A. Bassi. »From today's INTRAnet of things to a future INTERnet of things: a wireless- and mobility-related view«. In: *IEEE Wireless Communications, vol. 17, no. 6, pp. 44-51*. 2010.

[32] P. Maier and K. Nohl. *Low-Cost Chip Microprobing*. Accessed: 2022-03-08. 29th Chaos Communication Congress (29C3). URL: https://fahrplan.events.ccc.de/congress/2012/Fahrplan/attachments/2247_29C3-Dexter_Nohl-Low_Cost_Chip_Microprobing.pdf.

[33] A. Maiti and P. Schaumont. »The Impact of Aging on a Physical Unclonable Function«. In: *IEEE Transactions on Very Large Scale Integration Systems*. 2014.

[34] S. Manich, M. S. Wamser and G. Sigl. »Detection of probing attempts in secure ICs«. In: *IEEE International Symposium on Hardware-Oriented Security and Trust*. 2012.

[35] Martin et al. »Adaptive voltage scaling by in-situ delay monitoring for an image processing circuit«. In: *2012 IEEE 15th International Symposium on Design and Diagnostics of Electronic Circuits Systems (DDECS)*. 2012, pp. 205–208. DOI: 10.1109/DDECS.2012.6219058.

[36] Dominik Merli et al. »Semi-Invasive EM Attack on FPGA RO PUFs and Countermeasures«. In: *Proceedings of the Workshop on Embedded Systems Security*. WESS '11. Taipei, Taiwan: Association for Computing Machinery, 2011. ISBN: 9781450308199. DOI: 10.1145/2072274.2072276. URL: https://doi.org/10.1145/2072274.2072276.

[37] Seyed Hamidreza Moghadas and Georg Fischer. »Approaching bandwidth reliability of a CMOS highly dense physical unclonable function«. In: *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. 2017, pp. 160–165. DOI: 10.1109/IEMCON.2017.8117178.

[38] Seyed Hamidreza Moghadas and Georg Fischer. »Robust IoT communication physical layer concept with improved physical unclonable function«. In: *2017 IEEE Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics (PrimeAsia)*. 2017, pp. 97–100. DOI: 10.1109/PRIMEASIA.2017.8280373.

[39] Seyed Hamidreza Moghadas and Michael Pehl. »ROPAD: A Fully Digital Highly Predictive Ring Oscillator Probing Attempt Detector«. In: *2020 57th ACM/IEEE Design Automation Conference (DAC)*. 2020, pp. 1–6. DOI: 10.1109/DAC18072.2020.9218546.

[40] Seyed Hamidreza Moghadas, Michael Pehl and Georg Sigl. »ROPAD $^+$: Enhancing the Digital Ring Oscillator Probing Attempt Detector for Protecting Irregular Data Buses«. In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* (2022), pp. 1–12. DOI: 10.1109/TVLSI.2022.3191471.

[41] M. Mustapa and M. Niamat. »Temperature, Voltage, and Aging Effects in Ring Oscillator Physical Unclonable Function«. In: *2015 IEEE 17th International Conference on High Performance Computing and Communications, New York, NY, 2015, pp.1699-1702.*

[42] B. C. Paul. »Temporal Performance Degradation under NBTI: Estimation and Design for Improved Reliability of Nanoscale Circuit«. In: *Proc. ACM/IEEE Design Automation Conf. (DAC)*. 2006.

[43] T. Sakurai and A. R. Newton. »Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas«. In: *IEEE Journal of Solid-State Circuits* 25.2 (1990).

[44] Soumyajit Sarkar and Arijit Basu. »Comparison of various Edge Detection Techniques for maximum data hiding using LSB Algorithm«. In: *International Journal of Computer Science and Information Technologies 5.3 (2014): 4722-4727.*

[45] G. E. Suh and S. Devadas. »Physical unclonable functions for device authentication and secret key generation«. In: *Proc. 44th ACM IEEE Design Autom Conf.* 2007.

[46] S.V.Kumar. »Adaptive Techniques for Overcoming Performance Degradation due to Aging in Digital Circuits«. In: *Proc. IEEE ASP-DAC*. Jan. 2009.

[47] C. Tarnovsky. *Deconstructing a Secure Processor*. Accessed: 2022-04-28. Blackhat DC. URL: `https://www.blackhat.com/presentations/bh-dc-08/Tarnovsky/Presentation/bh-dc-08-tarnovsky.pdf`.

[48] et al. V. Reddy. »The Impact of NBTI on the Performance of Combinational and Sequential Circuits«. In: *Proc. ACM/IEEE Design Automation Conf. (DAC)*. 2007.

[49] H. Wang et al. »Probing Attacks on Integrated Circuits: Challenges and Research Opportunities«. In: *IEEE Design Test* 34.5 (2017).

[50] W. Wang. »Compact Modeling and Simulation of Circuit Reliability for 65nm CMOS Technology«. In: *IEEE Transactions on Device and Materials Reliability*. 2007.

[51] M. Weiner et al. »A Calibratable Detector for Invasive Attacks«. In: *IEEE Transactions on Very Large Scale Integration Systems* 27.5 (2019).

[52] M. Weiner et al. »The Low Area Probing Detector as a Countermeasure Against Invasive Attacks«. In: *IEEE Transactions on Very Large Scale Integration Systems* 26.2 (2018).

[53] Florian Wilde, Matthias Hiller and Michael Pehl. »Statistic-based security analysis of ring oscillator PUFs«. In: *2014 International Symposium on Integrated Circuits (ISIC)*. 2014, pp. 148–151. DOI: `10.1109/ISICIR.2014.7029528`.

[54] Martin Wirnshofer et al. »A variation-aware adaptive voltage scaling technique based on in-situ delay monitoring«. In: *14th IEEE International Symposium on Design and Diagnostics of Electronic Circuits and Systems*. 2011, pp. 261–266. DOI: `10.1109/DDECS.2011.5783090`.

[55] Martin Wirnshofer et al. »An energy-efficient supply voltage scheme using in-situ Pre-Error detection for on-the-fly voltage adaptation to PVT variations«. In: *2011 International Symposium on Integrated Circuits*. 2011, pp. 94–97. DOI: `10.1109/ISICir.2011.6131888`.

[56] R. Wittmann. »Miniaturization Problems in CMOS Technology: Investigation of Doping Profiles and Reliability«. In: *PhD Dissertation, TU Wien, 2007*. URL: `http://www.iue.tuwien.ac.at/phd/wittmann/node10.html`.

[57] R. Wittmann et al. »Impact of NBTI-driven parameter degradation on lifetime of a 90nm p-MOSFET«. In: *2005 IEEE International Integrated Reliability Workshop*. 2005, 4 pp. DOI: `10.1109/IRWS.2005.1609573`.

[58] L. Shang X. Wang and H. Yin. »Reliability concerns on time-to-digital converter due to bias temperature instability in nanometer era«. In: *2015 IEEE 11th International Conference on ASIC, Chengdu, 2015, pp. 1-4.*

[59] Teng Yang et al. »14.7 In-situ techniques for in-field sensing of NBTI degradation in an SRAM register file«. In: *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*. 2015, pp. 1–3. DOI: `10.1109/ISSCC.2015.7063027`.

[60] Teng Yang et al. »insitu and In-Field Technique for Monitoring and Decelerating NBTI in 6T-SRAM Register Files«. In: vol. 26. 11. 2018, pp. 2241–2253. DOI: `10.1109/TVLSI.2018.2856528`.

Lists

List of Figures

List of Tables