UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

“PHYSICAL UNCLONABLE FUNCTION”

Daniel Andrés Criollo Caisaguano

Profesor
Felice Crupi, PhD.

Trabajo de titulación de posgrado presentado como requisito
para la obtención del título de Máster en Nanoelectrónica

Quito, 10 diciembre 2019

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

“PHYSICAL UNCLONABLE FUNCTION”

Daniel Andrés Criollo Caisaguano

Felice Crupi

Prof. Felice Crupi, PhD.

Director de Trabajo de Titulación

Omar Aguirre, PhD.

Director de la Maestría en Nanoelectrónica

César Zambrano, PhD.

Decano del Colegio de Ciencias e Ingenierías

Hugo Burgos, PhD.

Decano del Colegio de Posgrados

Quito, 10 diciembre 2019

© Derechos de Autor

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma del estudiante: _______________

Nombre: Daniel Andrés Criollo Caisaguano

Código de estudiante: 203297

C. I.: 1718809666

Lugar, Fecha: Quito, 11 de diciembre 2019

DEDICACIÓN

El presente documento dedico a mis padres Luis Criollo y Rocío Caisaguano, por el amor incondicional, la comprensión, la paciencia, y sobre todo por creer en mí, por hacer que mi vida sea más hermosa y fructífera.

A mi hija quien es la mayor inspiración de vida.

DANIEL

AGRADECIMIENTO

El presente trabajo investigativo lo dedico principalmente a Dios, por darme fuerza para continuar este proceso de obtener uno de los anhelos más deseados.

A mis padres, por su amor, trabajo y sacrificio en todos estos años, gracias a ellos he logrado llegar hasta esta instancia de mi vida y por inculcarme a no temer las adversidades porque Dios está conmigo siempre.

A mis hermanos por estar siempre presentes, acompañándome y por todo ese aployo que nos brindan a lo largo de esta etapa de nuestras vidas.

Agradecer a mi tutor y a todos aquellos que nos abrieron las puertas y compartir su conocimiento. Al igual que la colaboración de Massimo Vatalaro.

DANIEL

RESUMEN

Este documento proporciona una introducción completa sobre el diseño de seguridad, aplicaciones de autentificación y generación de claves de una PUF (Physical Unclonable Function). La característica distintiva de los PUF es que no se puede crear una copia del circuito, ya que es imposible controlar las variaciones del proceso de fabricación. La autenticación de circuitos integrados y protocolos criptográficos seguros basados en hardware proporcionan un papel importante en la seguridad del hardware. La inserción de troyanos de hardware debido a la producción de circuitos integrados, los ataques de seguridad que requieren de acceso físico al dispositivo, se están volviendo más factibles dada la naturaleza dominante de la tecnología IOT. Las PUF son primitivas criptográficas innovadoras y fascinantes para aplicaciones de seguridad, estos dispositivos explotan las variaciones del proceso intrínsecas a los procesos de fabricación para generar una clave única para cada chip, como una especie de huella digital del dispositivo y generalmente son simples de implementar, pero difíciles para ser replicado.

El documento se divide en varios capítulos en los que los PUF se contextualizarán inicialmente dentro de la seguridad cibernética, analizando además de las posibles implicaciones de la aplicación, también los desafíos que deben enfrentarse durante su diseño. En el segundo capítulo se analizará el estado del arte, en la sección 3 se examinarán los resultados esperados, teniendo en cuenta los numerosos desafíos que el diseñador debe enfrentar durante el diseño de PUF y, finalmente, en el último capítulo, un resumen final.

Palabras Clave: Internet de las Cosas (IoT), Circuito Integrado (IC), Memoria estática de acceso aleatorio (SRAM), par desafío-respuesta (CRP), Funciones Físicas no clonables (PUFs).

ABSTRACT

The objective of the paper is to provide a complete introduction to the design of physical unclonable functions (PUF), devices used for applications such as low-cost authentication and generation of cryptographic keys. The distinctive feature of PUFs is that, being devices that take advantage of process variations, it is not possible to create a copy of the circuit that exactly replicates the behavior of the starting PUF, so it is intrinsically robust to any invasive physical attacks. Integrated circuit authentication and hardware encryption protocols play an important role in cyber security. The insertion of Trojans at the hardware level during the manufacturing of integrated circuits, requiring physical access to the device, is becoming a crucial aspect with the rapid development of the IoT. PUFs are fascinating innovative cryptographic primitives for security applications, these devices exploit the process variations intrinsic to the manufacturing processes to generate a unique key for each chip, as a sort of fingerprint of the device and are generally simple to implement but difficult to be replicated.

The paper is divided into several chapters in which the PUFs will initially be contextualized within cyber security, analyzing in addition to possible application implications, also the challenges that must be faced during their design. In the second chapter the state of the art will be analyzed, in section 3 the expected results will be examined, taking into consideration the numerous challenges that the designer must face during the PUF design and finally in the last chapter, a final summary.

Key Words: Internet of Things (IoT), Integrated Circuit (IC), Static Random - Access Memory (SRAM), Challenge - Response Pair (CRP), Physical Unclonable Functions (PUFs).

CONTENTS INDEX

TABLE INDEX

FIGURE INDEX

1 INTRODUCTION

The IoT known as the Internet of Things is a technology that has grown rapidly in recent years. It is used in various applications such as home automation, industrial production and many other computer applications. With the increase in the number of devices connected to each other, and the consequent increase in the amount of data transmitted, the need for greater data protection has also increased. When accessing a bank account, unlocking a mobile phone or any electronic device, an authentication phase is required for correct and secure access to the data. In order to avoid any type of attack, all information must be securely protected, which is why it is becoming increasingly important to develop different technologies that can efficiently solve any type of security and identification problem.

Different technologies of this type are destined to be implemented on FPGA, thanks to the continuous development of integrated circuits (IC) and intellectual property (IP), used in a wide range of applications. Generally, to guarantee the correct identification of a device, a key stored in a non-volatile memory is used, which is requested during the authentication phase. The problem occurs when information is stolen, for example through the bitstream phase of the FPGA, or by cloning the chip by building an identical chip created with each of its components, compromising the authenticity of the chip and accessing all the secret information. Furthermore, data protection requires circuits external to the chip that must be continuously powered, leading to non-negligible energy consumption.

A physical element can be hindered by software procedures that can clone the physical characteristics of each of the devices, increasing the number of possible attacks on the user and the network. The increase in the number of devices connected to each other must be supported by a strong increase in security, which leads to an increase in the search for data protection in hardware and software.

An alternative to the use of non-volatile memories (NVM) for storing the security key are the Physical Unclonable Functions (PUFs), which offer a safe, low-cost solution, characterized by low area occupancy and low consumption of power, ideal for applications such as low-cost authentication and generation of an encryption key. A PUF is considered a function that exploits the process variations intrinsic to the manufacturing processes to generate a unique key for each device, such as a sort of fingerprint of the chip. Thanks to this property it becomes extremely complicated or impossible to produce two identical circuits on the same wafer with the same electrical properties. In this way it is possible to uniquely identify each device [1][2].

The growing development of the Internet of Things (IoT) and of the number of devices connected to each other (by 2020 an estimated 40 billion devices are estimated) on the one hand we allow ourselves to get to do things that previously could not be done, for example today we use RFID technology for controlled access to homes or for example for the proper execution of banking transactions, on the other hand with the development of these denser networks of devices it is becoming very important to obtain security measures suitable for protection against cyber-attacks.

Generally, to obtain a unique identification of a device, a random key generated and stored in the non-volatile memory of the device during the design phase is used during the authentication phase. The disadvantages of this approach are related to the possibility that the chip is cloned, and consequently also all the delicate information contained therein, and the energy costs required for data protection due to an external circuitry that must always be powered. For this reason, in recent years there has been a great deal of research and development into technologies capable of allowing a reliable authentication phase without having to resort to storing the key in its memory. The Physical Unclonable Functions (PUFs), for example, are innovative cryptographic primitives able to exploit the process variations or the intrinsic fluctuations of some physical parameters to generate unique keys for each device as a sort of fingerprint of the chip. Over the years the study of these devices has led to different solutions including the use of random light scattering to exploit the intrinsic fluctuations of some geometric and physical parameters during chip manufacturing processes, increasing with technological scaling, resulting in variation of the electrical characteristics of devices manufactured on the same wafer, to generate ideally unique keys for each device [3][4].

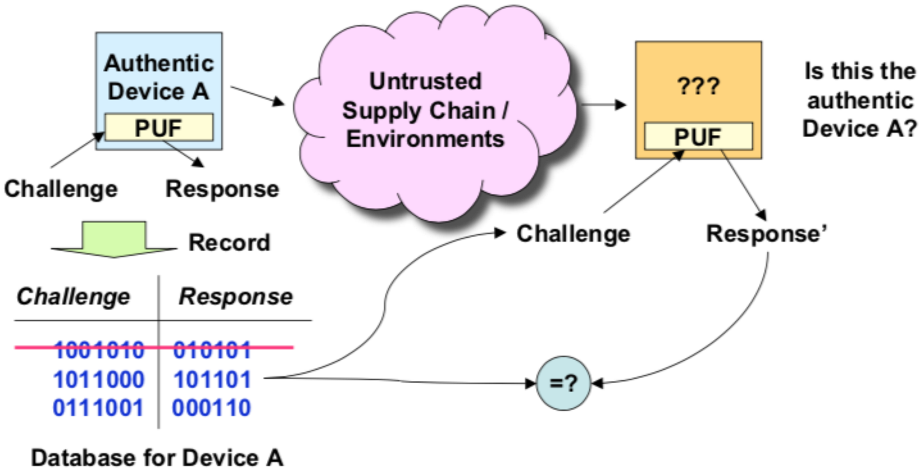
The properties that make PUFs attractive for IT security applications are the following:

- **Randomness,** which ensures the same number of ones and zeroes within the generated key, this to make probabilistic attacks from the outside more difficult.
- **Physical Unclonability,** which ensures the impossibility of creating identical PUFs, even if made on the same wafer, so even though they are structurally identical, if stimulated with the same challenge they will respond with different keys.
- **Unpredictability,** which ensures that process variations cannot be predicted through mathematical algorithms.
- **Uniqueness,** which ensures a unique behavior of each PUF.

1.1 PUF typologies and applications

As mentioned above, the mechanism on which the operation of a PUF is based is a challenge-response pairs, in which challenge means the way in which the PUF is stimulated and the response is the key it generates from the challenge. Precisely for this reason we can distinguish two macro types of PUF that differ according to the number of challenge-response pairs (CRPs) that they can generate.

There are **strong PUFs** characterized by a large number of challenge-response pairs whose main field of application is the low-cost authentication phase in which the authenticator stores in a database all the possible CRPs associated with a device and to verify its authenticity before granting him access to the data stimulates him with a challenge and verifies if the response of the PUF, integrated on that device, to that challenge is contained in the database [5].



Challenge	*Response*
~~1001010~~	~~010101~~
1011000	101101
0111001	000110

Figure 1. Overview of authentication in a PUF

To ensure greater immunity to any attacks on the database, since it is the only device that needs to memorize all the keys, and to probabilistic attacks, the authenticator deletes each CRP after use so that each of them can be used only once [6].

The second type of PUF is the **Weak PUF**, characterized by a relatively small number of challenge-response pairs and used mostly for the generation of cryptographic keys during the data encoding process, to achieve greater security during transmission, in which the exchange of keys can take place, for example, in a symmetrical or asymmetrical manner [4].

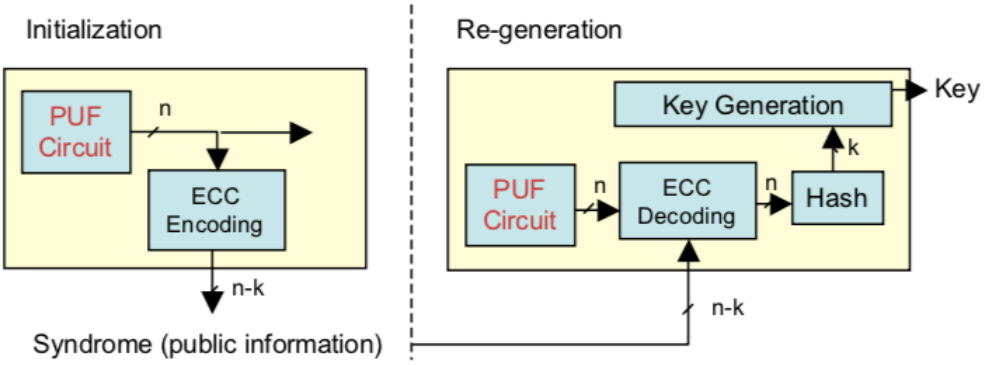


Figure 2. Generation of cryptographic keys with PUF

1.2 PUF architectures

As mentioned above, there are various types of PUF, from optical to silicon-based ones. Below is an overview of the state of the art.

1.3 Optical PUF

Long before knowing the meaning of PUF, a system based on random optical reflection patterns, known as reflective particle labels, has been proposed. One of the applications in which we can use these types of labels is in the arms control treaties (TLI). The main objective is to provide a single verification for each device and prevent them from falling into ambiguous hands due to espionage. Among the requirements that must be met is the impossibility of falsification. This means that the labeling system must be impossible to clone and must not reveal secret information without prior consent. The physical dimensions and energy requirements of the label must be minimal, and it must also be reliable, that is, it must have a low false alarm rate and it must be simple and economical to implement. In addition, the device label shall provide all necessary information about the reading process so that it can be read objectively.

Optical PUFs are devices designated as one-way physical function. For the construction initially a protocol of reflective particles was proposed, based on random optical reflection patterns. The design of the device can be seen in Figure 3, and consists of a transparent optical medium known as an "optical token" which is filled with a large number of light-scattering particles and at the time when a laser beam illuminates the optical medium, a point is produced with a random and unique pattern, taking advantage of the random scattering of light, that is the randomness in this type of PUF is manifested due to the random distribution of the particles of light dispersion in the optical medium during the manufacturing process. The resulting spot is recorded and encoded in a bit string that represents the PUF output. In this type of PUF the input challenge can be the angle in which the laser beam that hits the optical medium is fixed [7].

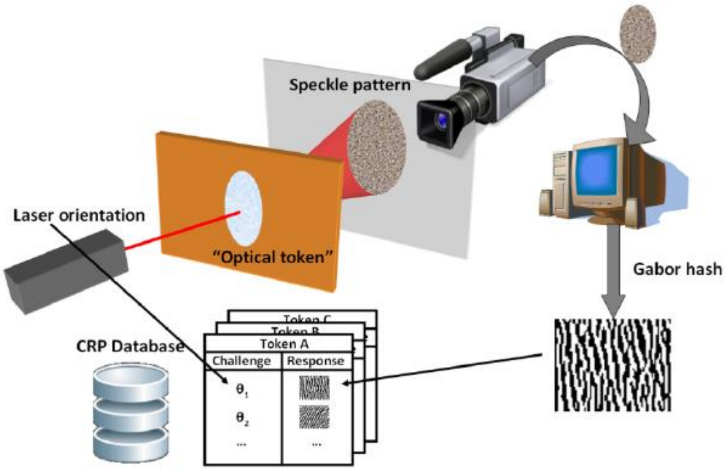


Figure 3. Optical PUF

One of the typical examples that can be found are credit cards that contain a three-dimensional inhomogeneous microstructure with features that resize the wavelength to visible light. In Figure 4. We can observe the principle of operation, in which the laser beam must initially pass through the token and due to the random distribution of light scattering particles and how they interact with the laser, many spots are produced in the output image. This model of speckles generated is used to derive a unique identifier for the structure. This process be a physical hashing of the complex 3D structure up to a fixed-length key and we need a conversion process for which we use the Gabor Hash transformation to learn about mottled models and reduce them to a bit string [8].

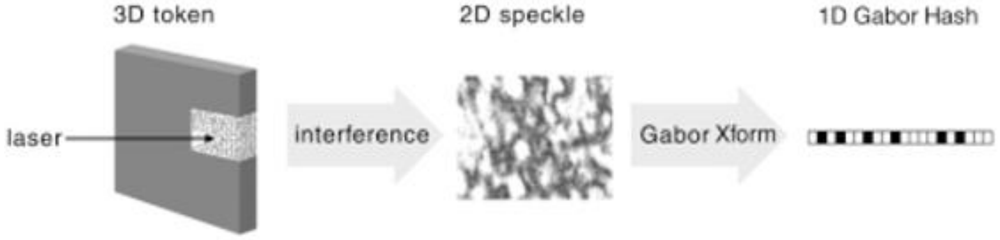
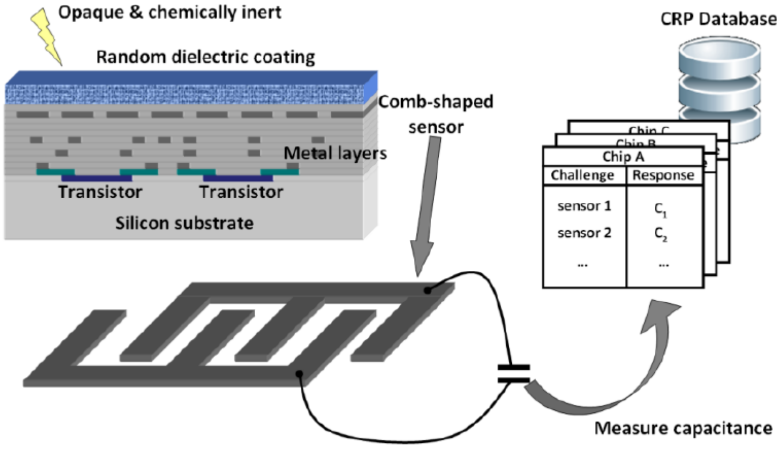


Figure 4. Basic structure on Physical One-Way Functions

1.4 Coating PUF

A PUF coating can be made on the surface of an integrated circuit and consists of covering the circuit itself with a protective coating. Then, taking into consideration a network of wires that typically comes in the form of a comb, and covering it through an opaque and doped material with dielectric particles, whose size, shape and permittivity change from chip to chip due to variations during deposition process on an integrated circuit. These particles are a mixture of TiO_2 and TiN in an aluminophosphate matrix that ensures light absorption properties ranging from infrared light to ultraviolet light. The intrinsic randomness of the dielectric particle deposition process means that when measuring the PUF coating from the outside, different capacitance values will be obtained since the measurements are very sensitive to their position. This randomness can be used to obtain a unique identification for each device on which the PUF coatings are made. In addition, these circuits provide a great protection against physical attacks, since the manipulation of the coating causes changes in the electrical properties, changing the value of the output capacitance. The positioning of these PUFs on the top layer of an integrated circuit protects the underlying circuits against attacks by an attacker, for example by reverse engineering. In fact, when an attempt is made to remove part of the covering, the capacity between the wires is destined to change and the original unique identifier is compromised. An application field, for example, is to build a non-clonable RFID tag through these PUF coatings [9].



Challenge	Response
sensor 1	C_1
sensor 2	C_2
...	...

Figure 5. Basic operation of a PUF Coating

1.5 Spintronic PUF

The next generation of PUFs begins to be implemented using nano-electronic and emerging quantum devices. For example, nanotechnologies such as the STT-MRAM (Magnetic Random Access Memory Spin-Torque) showed promising features for logic and memory applications thanks to greater energy efficiency, greater non-volatility and greater integration density. The non-linear dynamics of the regions of magnetic domain (DW), regions of ferromagnetic material characterized by a uniform magnetization that is the magnetic moments of the atoms present in the domain are aligned and have the same direction, can be exploited for the safety of the hardware. Spintronic circuits can be a complement to the silicon substrate to complement existing CMOS-based safety and reliability features.

Numerous experiments on magnetic tunnel junction (MTJ) and domino wall memory (DWM) have been conducted, obtaining a modulation effect induced by the dynamic magnetization current. The relationship between the injected current and the magnetization generates some spin-transfer-torque (STT) mechanisms that are excellent sources of entropy in the magnet. In Figure 6 we can observe each of the sources of entropy and randomness that these spintronic systems like MJT and DWM have [10].

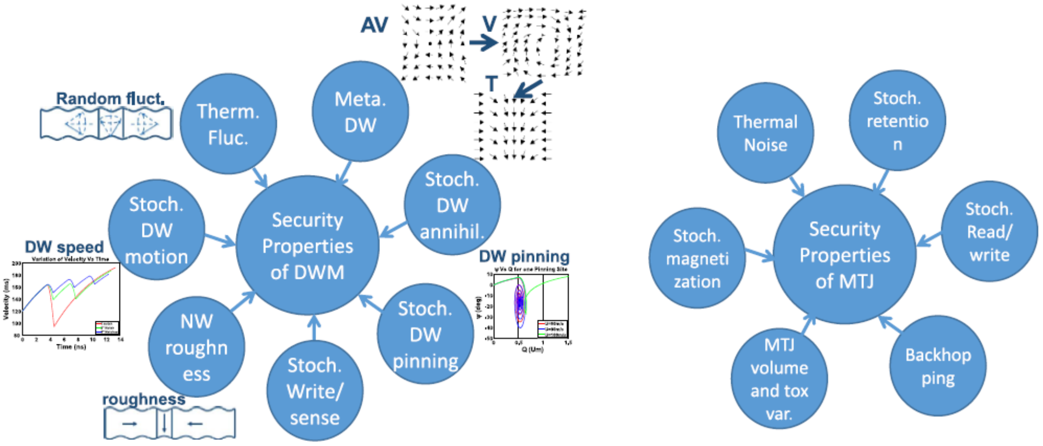


Figure 6. Sources of entropy and randomness in spintronic systems such as DWM and MTJ

It is known that spintronics technology is an excellent option for hardware security design, but it should also be considered that they can be easy prey against malicious attacks since they are very sensitive to parameters such as temperature and magnetic field. This makes us understand that new challenges come into play, as far as safety is concerned, during the design, which were not present in volatile memories like SRAM and DRAM.

For example, the MTJ memories are composed of two layers of ferromagnetic material separated by a thin tunnel barrier (typically MgO). The magnetic orientation of one of the two ferromagnetic layers is fixed, while the orientation of the other determines the state of the memory. Generally, an MTJ has two stable states, related to the magnetic orientation of the two levels, parallel (P) and anti-parallel (AP) and to these states are associated two levels of resistance ($R_P < R_{AP}$). The advantages of this type of memory are linked to the high integration density, to the low standby power and to the speed with which data writing and reading operations are performed. The change in the magnetic orientation of the free level occurs by imposing the right current in the branch, but what makes the MTJ suitable to their use for security applications such as the realization of a PUF, is linked to the fact that, given a current, not it is said that all cells are written in the same period of time. This means that, under certain conditions, the writing of the data occurs probabilistically. For example, we could use a current with a certain intensity, to write on a certain number of cells, for a fixed period, and then see in which cells the value of the stored data has changed.

The domino wall memory (DWM) is a promising technology thanks to the performances offered in terms of number of stored bits per cell, standby power consumption, non-volatility of the stored data, good resistance and low retention. A DWM is composed of a reading head, a writing head (very similar to MTJ) and a magnetic nano-cable, which keeps the bits in terms of magnetic polarity. DWs can inject charge current as they move back and forth, from right to left, making the nanowire function as a shift register. The safety properties of a DWM have microscopic and macroscopic properties.

In Table 1, we can see a comparison between the parameters of the safety primitives and the characteristics offered by spintronic. Details are provided as follows.

Safety primitive	Key Requirements	Features offered by spintronics
Recycling sensor	Low process variation, high sensitivity to use	DW nucleation
PUF	High process variation, non-linearity	Stochastic DW movement in raw nanowire, non-linearity
TRNG	High entropy	Noise sensitivity of magnetization, stochastic dynamics
Cryptography	Recursive displacement, multiplication, addition	Calculation based on shifts
Miscellaneous	Sensitivity to environmental parameters	Sensitivity to the magnetic field, temperature

Table 1. Comparison between the parameters of the safety primitives and the characteristics offered by spintronic

1.6 Silicon PUF

As regards the PUFs realized on silicon, they exploit the process variations, intrinsic to the manufacturing processes, linked to the variation of geometric parameters such as the channel length, the thickness of the oxide or electrical parameters such as the doping level to generate a unique key. Generally, the block diagram of a PUF can be represented in the following way [11].

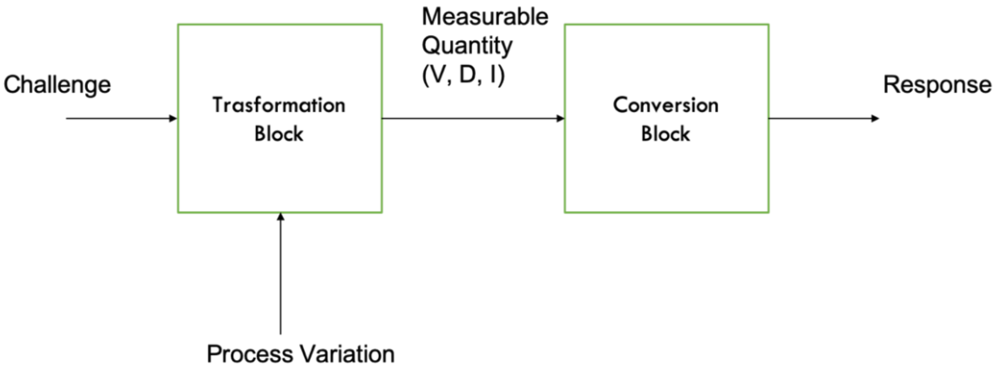


Figure 7. Generic architecture for a PUF

Inside we find a transformation block that has the task, when stimulated by a challenge, of transforming the process variations into a measurable quantity such as a voltage, a delay or a current and a conversion block whose task is to convert the information exiting the transformation block into binary. This makes us understand that there are three macro categories of silicon PUF:

- Delay based PUF
- Voltage based PUF
- Current based PUF

Before describing some type of PUF belonging to these subclasses, we will briefly describe what is meant by mismatch.

1.7 Mismatch

Given a circuit we can attribute the variability to two main aspects, such as variations in operating conditions, such as the variation of the supply voltage or temperature, which cause a degradation of the electrical performances, and the process variations that represent the physical imperfections of a device, linked to the fact that with continuous technological scaling it is increasingly difficult to create devices with extremely precise dimensions.

With the technological scaling and the realization of smaller and smaller devices, it becomes increasingly difficult to have full control, during the manufacturing process, over the dimensions of the same. Usually these fluctuations are not well seen by the designers because they are the main cause of the malfunctioning of the circuits and from the deviation of the results obtained after the layout compared to those expected before the layout for this reason it is important during the design phase, generally, to try take these variations into account, using appropriate models and appropriate simulations, to obtain results, during the analysis phase, close to those measured. On the other hand, however, these variations are a fingerprint of the chip itself because they provide it with a sort of identifier different from all the others and can be used to create these circuits whose purpose is to have some features that make them unique behavior. This makes us understand that one of the main advantages of PUFs is that they are simple to make but very difficult to replicate, precisely because process variations make each chip unique.

Generally the properties of a device can vary from wafer to wafer (inter-die variations), causing a type of transistor to have greater force than the other, or between transistors belonging to the same manufacturing process (intra-die variations) such as those related to the geometry of the devices that have a direct impact on one of the parameters that most suffer from it, which is the threshold voltage. At the device level the main sources of variability are related to the geometry (the most important are related to the variation of the thickness of the oxide, T_{OX}, and the variation in the size of the channel is in length, $L_{Channel}$, that in width, $W_{Channel}$, that are caused by the lithography processes) and to the material of the same (Regarding the material the major source of variability is linked to the variation of the doping levels of the devices, which is the main cause of the different behavior from the point of view of performance between nMOSFET and pMOSFET, and the depth of the diffusions).

In general, during a manufacturing process the sources of variability are not only linked to the devices but also to the interconnections (this shows that it is important to also do a post layout analysis) and also in this case the main sources of variability are related to geometry (the main sources of variability are related to the variation of the width and the thickness of the lines and to the variation of the dielectric height) and to the materials (the most important sources of variability are linked to the variation of the metal resistivity, to the variation of the dielectric constant , due to the deposition process, and finally to the variation of the contact resistance of the interconnections) [12].

Continuing with technological scaling, it becomes increasingly difficult to keep manufacturing processes under control, which makes it difficult to have accurate estimates of circuit performance. The following is a trend that explains how the variability on the threshold voltage of the devices and on the performance of the circuits has increased over the years.

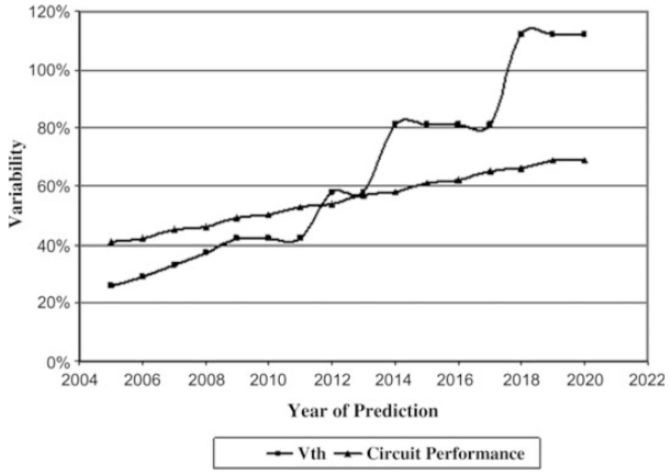


Figure 8. Impact of variability on the electrical parameters of the circuits

If on the one hand these results create problems for designers who must always take into consideration the worst case so that everything works correctly, on the other hand they are results that favor the design of PUF.

1.8 Delay-based PUFs

The delay-based PUF, when stimulated by a challenge, transform the process variations into a delay that is subsequently converted into binary. Below are some examples of PUFs belonging to this typology.

1.8.1 Arbiter PUFs

The first PUF taken under examination is the following.

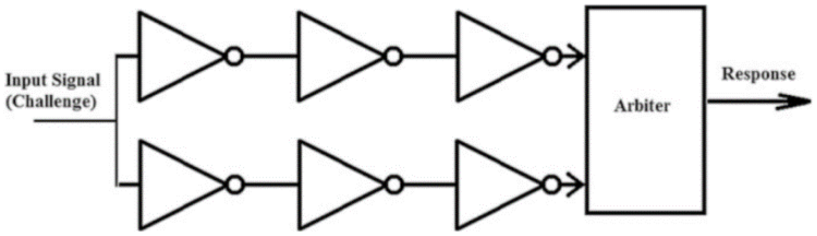


Figure 9. Architecture of a PUF Arbiter

As can be seen from the figure above, the circuit consists of two identical digital paths, therefore with the same nominal delay, and in an "arbitrator" whose task is to establish in which of the two paths the signal propagates faster, this is because due to process variations and how these are reflected on the transistor threshold voltage, when an input signal is applied it propagates with different delays in the two paths and therefore it will come out at two different times and the referee's task is to establish which of the two paths has "won" the race.

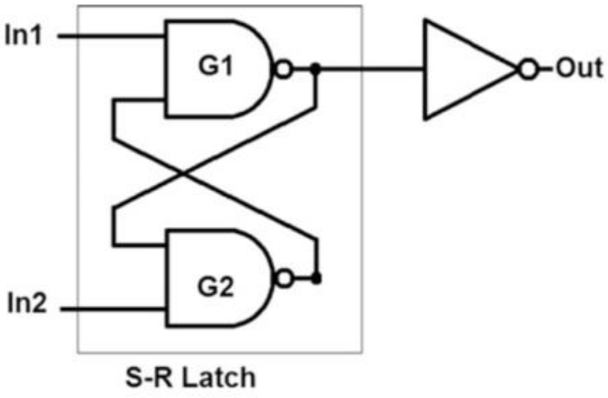


Figure 10. Arbiter architecture

For example, exiting the arbitration circuit, which could be a simple S-R latch as shown in the figure above, we will have 1 or 0 depending on which path the race wins. In case you want to create a system capable of generating many challenge-response pairs, you could use multiplexers along the various paths and make the paths participating in the race consist of many smaller paths selected by the incoming challenge , as shown in the figure below [13].

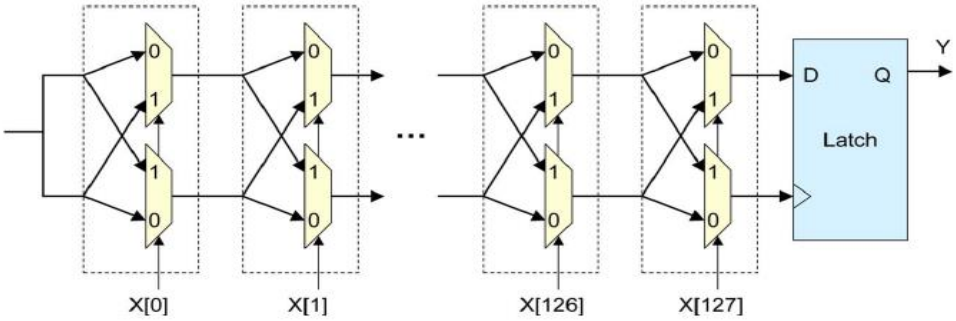


Figure 11. PUF Arbiter delay circuit. The circuit creates two delay paths with the same design length for each X input and produces an Y output depending on which path is faster

In this case, for example, each bit of the challenge drives a multiplexer whose task is to select one of the two paths connected to its inputs. For example, the challenges here could be $S = 00$ Or $S = 01$. Generally, the maximum number of CRPs of a PUF arbiter composed by k multiplexer is:

$$CRPs = 2^k$$

The main problem is related to the error introduced, for example, by the problem of the metastability that afflicts the S-R latch, in the case in which it was used as an arbitrator.

1.8.2 Ring Oscillator PUFs

The information coming out of this type of PUF is a frequency information, more precisely this type of PUF are generally composed, as shown in the figure below, by two multiplexers, two meters, a comparator and by N ring oscillator [8]. A ring oscillator is a circuit made up of a cascade of inverters, if the number of inverters in the path is odd, retroacting the output of the last stage of inverter connecting it to the input of the first stage, we will obtain that the output voltage will oscillate and the number of oscillations per second will depend on how the process variations affect the delay of the single path.

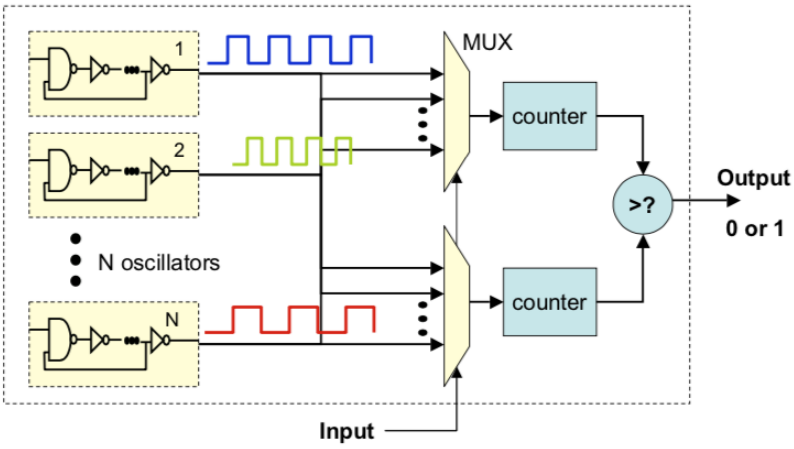


Figure 12. Ring Oscillator PUF

The purpose of the multiplexers is to create a multibit structure, in fact the multiplexer selector corresponds to the incoming challenge and will select the two ring oscillators to compare, one for multiplexers. On exit from the latter, the meters will count the number of oscillations observed by the relative ring oscillator in a predetermined time interval. Depending on which of the two counters gives a higher number at the output, the PUF output will settle at 1 or 0. Compared to the previous solution we solve the problem of metastability. The maximum number of CRPs of this type of PUF composed of N ring oscillator is:

$$CRPs = \frac{N\,(N-1)}{2}$$

Compared to the previous solution, it occupies a larger area and consumes more power. However, it reduces the number of errors associated with the metastability point caused by the S-R latch. Both solutions described are vulnerable to attacks with machine learning.

1.9 Current-Based PUFs

This type of PUF transforms the process variations into information in current, considering that the exponential dependence of the leakage current with the threshold voltage of the transistors emphasizes the variations due to the mismatch, to convert it later into binary.

1.9.1 Current-Based PUFs Using Transistor Arrays

As mentioned above, the idea is to exploit the exponential dependence of the sub-threshold currents with the transistor threshold voltage to increase the unpredictability of the PUFs behavior. The general architecture is as follows.

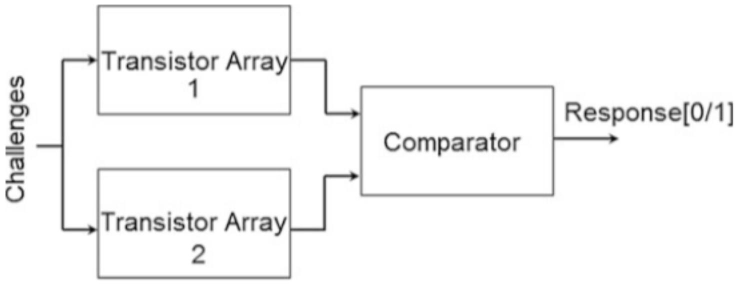


Figure 13. Architecture of a current based PUF using a transistor matrix

The idea behind this structure follows the previous proposals a little and consists in sending the same signal, challenge, to the two arrays and transforming the potential difference in output, due to process variations, into binary through a comparator.

The number of challenge-response pairs that this type of PUF can generate depends on how the transistor arrays are structured.

$$CRPs = 2^{kn}$$

Where k is the number of columns while n is the number of rows in the array. A possible circuit solution for realizing the transistor arrays is shown in the figure below, in which each cell is composed of 4 transistors, typically using two NMOS and two PMOS per cell. Where k is the number of columns while n is the number of rows in the array. A possible circuit solution for realizing the transistor arrays is shown in the figure below, in which each cell is composed of 4 transistors, typically using two NMOS and two PMOS per cell. Forward as regards the parallel transistors one of the two is typically smaller, stochastic transistor (for example N11x) to emphasize the impact of process variations on the transistor threshold voltage while the other is called switch transistor (N11) and serves to remove the impact of the stochastic transistor from the network during its ON state, or to include it during the OFF state. Furthermore, in each array each cell has its dual, this means that if a high value is received at the input it will turn on NMOS of the green cell in the figure and the PMOS of the red cell in the figure, this with regard to the switch transistors, as far as regards the stochastic transistors the PMOS of the green cell and the NMOS of the red cell will be included in the network. The value of the output voltage of the single array will depend on the resistive path created and therefore on the impact of the process variations on the threshold voltage of the stochastic transistors. While the potential difference in input to the comparator represents the difference in output potential between the two arrays [14].

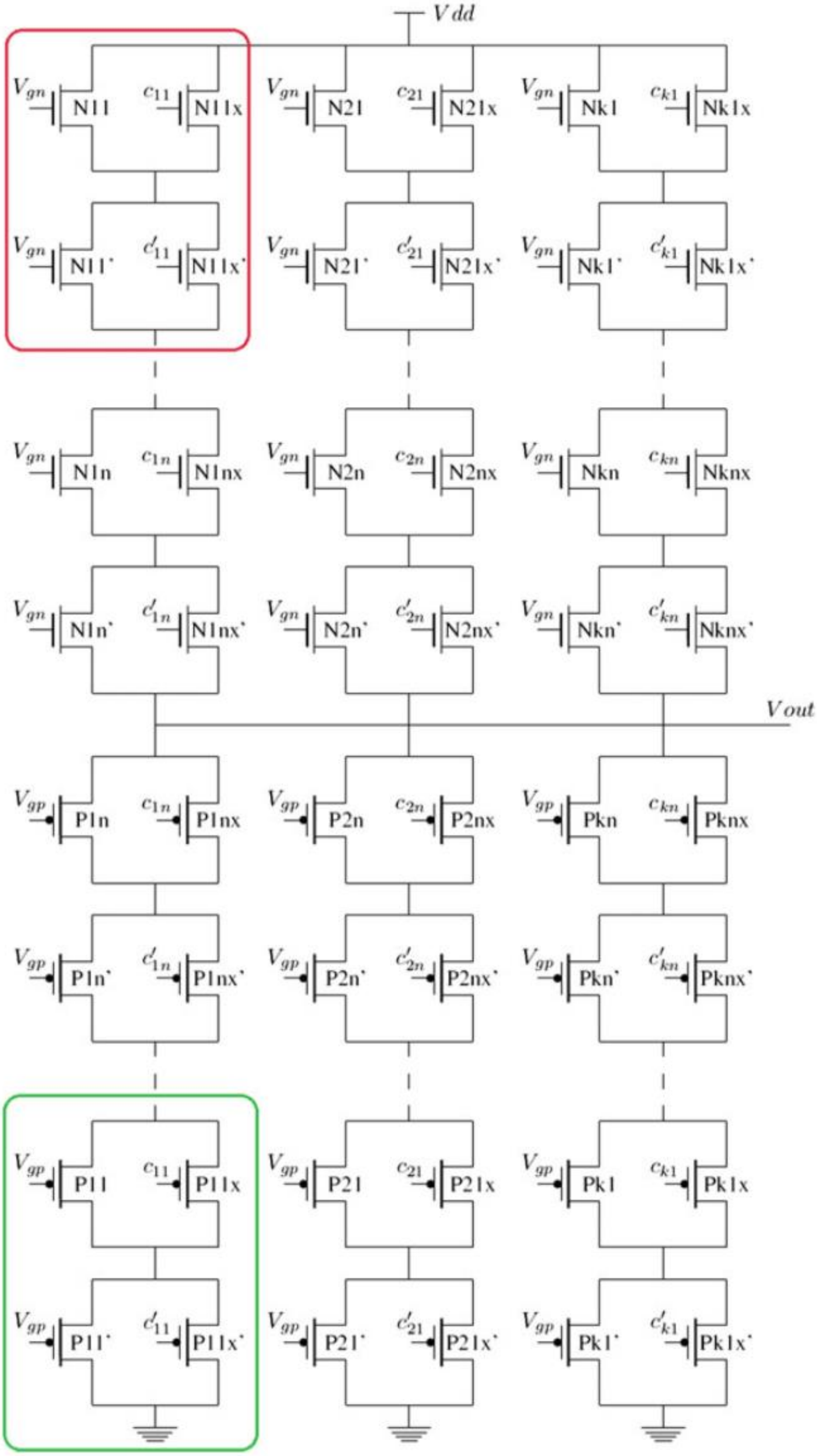

$V dd$
V_{gn} N11 c_{11} N11x V_{gn} N21 c_{21} N21x V_{gn} Nk1 c_{k1} Nk1x
V_{gn} N11' c'_{11} N11x' V_{gn} N21' c'_{21} N21x' V_{gn} Nk1' c'_{k1} Nk1x'
V_{gn} N1n c_{1n} N1nx V_{gn} N2n c_{2n} N2nx V_{gn} Nkn c_{kn} Nknx
V_{gn} N1n' c'_{1n} N1nx' V_{gn} N2n' c'_{2n} N2nx' V_{gn} Nkn' c'_{kn} Nknx'
$V out$
V_{gp} P1n c_{1n} P1nx V_{gp} P2n c_{2n} P2nx V_{gp} Pkn c_{kn} Pknx
V_{gp} P1n' c'_{1n} P1nx' V_{gp} P2n' c'_{2n} P2nx' V_{gp} Pkn' c'_{kn} Pknx'
V_{gp} P11 c_{11} P11x V_{gp} P21 c_{21} P21x V_{gp} Pk1 c_{k1} Pk1x
V_{gp} P11' c'_{11} P11x' V_{gp} P21' c'_{21} P21x' V_{gp} Pk1' c'_{k1} Pk1x'

Figure 14. Structure of a matrix of transistors

1.9.2 Current Based PUFs Using Dynamic Random Access Memories

Generally, the cell of a dynamic random access memory (DRAM) is composed of an access transistor and a capacitor whose task is to store information and preserve it. As we know, however, this type of cell needs auxiliary circuitry that refreshes the data. This is because due to the leakage currents that afflict the transistor when the potential of the terminal to which the capacitance is connected and that to which the bitline is connected is different and this means that there is a current that discharges the value stored in the cell to mass. If for memory applications this current is harmful, for applications such as PUF this current can be exploited. The strategy adopted takes into consideration that the intensity of this current has an exponential dependence with the threshold voltage which is the parameter on which the process variations are most affected. This means that, once the high value has been initialized in the cell, the time it takes for a cell to discharge depends on the intensity of the current, so different cells will discharge with different times. Estimated the average discharge value of a cell then you can initialize the cells to a high logical value, then turn them off and see, after waiting for the necessary time, how many of these cells have been discharged and how many have not. Obviously about the cells that will be used to realize the PUF it is necessary to deactivate the data refresh.

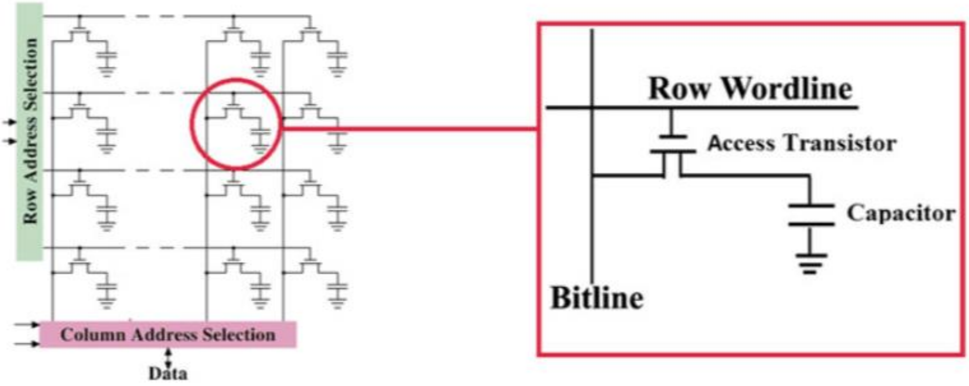


Figure 15. Architecture of a current-based PUF based on a DRAM memory cell

As for the number of obtainable CRPs, considering a DRAM with reserved regions in which N decay periods can be exploited, then:

$$CRP = R \, x \, N$$

It is very important to establish a suitable time frame before reading the various cells, because if you choose too small periods of time, there is a risk that in none of the cells will the logical value be changed, while if you wait too long periods of time then you risk that all the cells have already changed their logical state. One of the main advantages of this type is the fact that for its realization it does not require additional area occupation, because the DRAM cells already present can be exploited.

1.10 Voltage-Based PUFs

This type of PUF generally uses the metastability points of memory elements to generate a voltage information subsequently converted into binary.

1.10.1 SRAM PUFs

The use of static random access memory (SRAM) for generating a key was initially intended for FPGAs so that it could encode the bitstream, before storing it in an external memory, to protect it from possible attacks. As shown in the figure below, the structure of the cell of a 6T SRAM is composed of two feedback inverters to ensure that the direct and the denied output node are always connected either to the power supply or to ground, to avoid the problem of loss of data due to leakage currents. Obviously, with respect to DRAMs, the greater robustness, from this point of view, leads to an increase in area occupation, because it goes from having a transistor and a capacitor per cell with six transistors, taking into consideration also those of access, per cell.

To access the cell it is necessary to use the relative wordline and write the data by correctly managing the two bitlines, in this way the two inverters take care of supplying the nodes Q and Q' a low resistance connection to earth or to the power supply depending on the data stored, thus avoiding that the data can be lost.

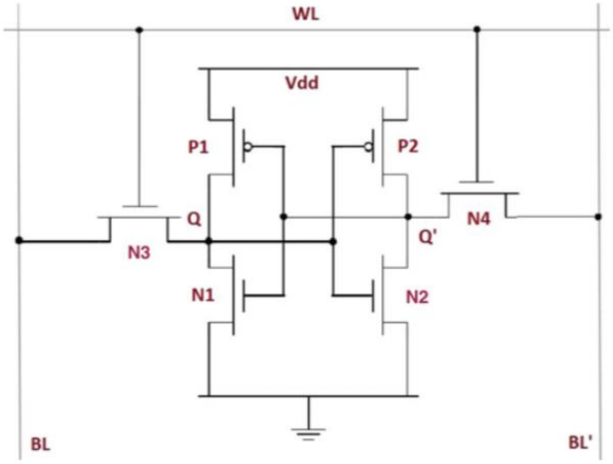


Figure 16. Architecture of a 6T SRAM memory cell

As shown in the figure below, a SRAM cell has three working points. Two stable and one metastable. In the two stable work points you find yourself once the data is stored, depending on the data itself [15].

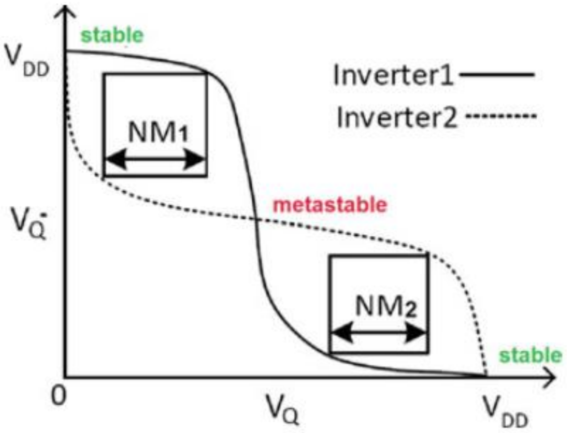


Figure 17. Noise margins of a SRAM memory cell and the work points

The phenomenon that we intend to exploit is the metastability point in which the RAM is working when it is turned on. At this point the data stored is neither one nor zero. What happens is that in the absence of mismatches between the transients, the cell could continue to work at this point, but due to process variations we find ourselves working with inverters that are not perfectly balanced, is one of the two could push the node that it drives more towards mass or towards the power supply and this would involve a forced charge or discharge of the node, triggering the positive feedback, writing a "random" datum and bringing the cell to work in one of the two stable points. The idea is to exploit the fact that when the cell is working at the point of metastability it is not possible to predict in which of the two states it will evolve, because it is linked to process variations. For the purposes of the operation of a PUF we could feed the cells destined for this purpose and read after a short period of time in which point of work the memory is working. Also, in this case one of the main advantages is the fact that an existing memory can be used without having to use additional circuitry. As for the number of challenge response pairs that can be generated, it depends on how many cells are allocated for this function and the possible challenges could be the memory addresses themselves.

1.10.2 Latch Based PUFs

Also, in this case we want to exploit the metastability point that afflicts the memory elements when they are turned on. The architecture is that of a latch, as shown in the figure below.

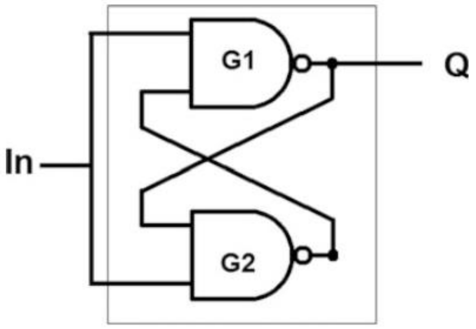


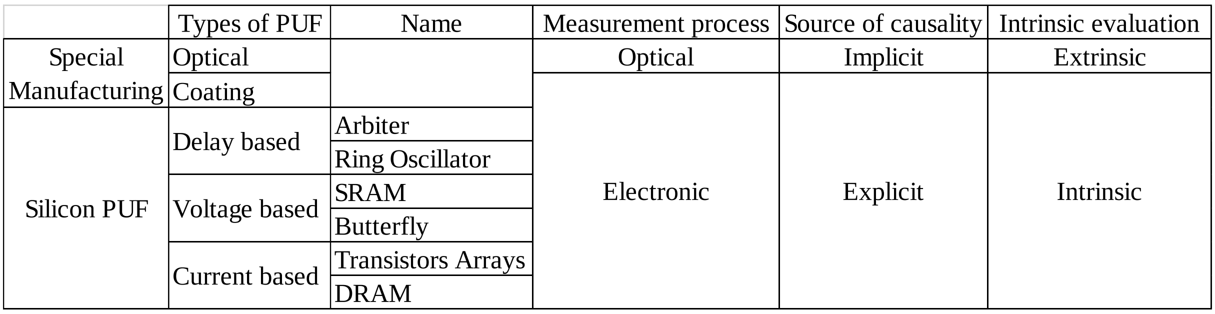
Figure 18. PUF latch based architecture

To force the metastability point, for example, we can start from the situation in which we have a zero in input and we have a one in output. If subsequently the logic state of the input is changed then the latch would enter a state of metastability, and the stable value of the output would depend also in this case on how it affects the process variations on the transistors. The main problem with this solution is the impossibility of accurately predicting when the latch will enter its metastability state and the amount of time to wait before it can read the outgoing data [16].

1.11 Comparison

This section intends to offer a comparison between the different metrics and characteristics of the PUF circuits, because they are subject to environmental variations such as temperature, supply voltage and electromagnetic interference that can affect their performance. The main objective of these circuits is to improve the security of electronic devices against malicious attacks and to obtain unique authentication and authorization for each of the devices. For this, the following summary is proposed.

The way of classifying the concepts one has on a PUF depends very much on how to measure the origin of the variation in each device. So, if you want to get a signature in a PUF circuit, you are referring to different processes. One of these is the devices that interact with an electronic signal and another is the devices that have the task of examining the effects on light or on different optical processes. The simplest way to authenticate an electronic circuit is through electronic characterization thanks to its ease of integration. Another of the main parameters is derived from the randomness or variation of the device. There are two types of random sources, one implicitly and the other explicitly. The difference between explicit randomness lies in the way in which CMOS components can be added without the need for further fabrication (electronic PUF), which is totally the opposite of explicit randomness because they are already part of the typical variation processes. If we add a dielectric for the sole purpose of detecting the fingerprints of a PUF, we are increasing the steps for manufacturing the device, in order to change the category to the explicit one. The implicit randomization sources have the advantage of low costs, since the components would not be added in the manufacture of the device. Explicit random sources offer the advantage of freely choosing a random source for the benefit of performance or increasing the difficulty of cloning a device, taking advantage of production to make them much smaller. Together with the sources of randomness we can intrinsically evaluate PUF devices, being able to describe that a PUF is considered intrinsic if its randomness is of implicit origin. Furthermore, the evaluation processing is performed under the supervision of electronic mechanisms. Table 2 shows different PUFs based on their types and their specific characteristics [17].



	Types of PUF	Name	Measurement process	Source of causality	Intrinsic evaluation
Special Manufacturing	Optical		Optical	Implicit	Extrinsic
	Coating		Electronic	Explicit	Intrinsic
Silicon PUF	Delay based	Arbiter			
		Ring Oscillator			
	Voltage based	SRAM			
		Butterfly			
	Current based	Transistors Arrays			
		DRAM			

Table 2. Comparison between different types of PUF

2 Figures of merit and comparison metrics

When designing PUFs it is important to consider some crucial aspects. More precisely, it is important to create a design that has the following characteristics:

- Very often the portion of circuitry intended for PUF is an additional circuitry and for this reason it is important that the area occupancy and consumption are contained.
- It is important that the single cell is strongly bistable, to ensure greater immunity to external interference that could cause malfunctions or errors during the authentication phase itself.
- It is important that the PUF response is stable to changes in operating conditions, such as supply voltage or operating temperature. This is because the database is filled by the authenticator with a series of CRPs obtained under certain operating conditions (for example at room temperature and at nominal voltage). If during the authentication phase the temperature or the supply voltage deviate from the nominal ones it is important, in order not to incur in a wrong authentication, that the response of the PUF is the same, or at least that the number of possible bits that are flipped is contained in a relatively low number so that it can be corrected with appropriate error correction circuits.

2.1 Comparison metrics

It therefore becomes important to use metrics to compare the results obtained from different types of PUF. The metrics analyzed in this chapter to assess the quality of the design are as follows:

- Uniqueness
- Reliability
- Uniformity
- Handling Strength

Before proceeding with the description of these metrics it is good to introduce the concept of Hamming distance and Hamming weight [11][16].

- **Hamming distance:** Give two words a, b of length n, the Hamming distance $\boldsymbol{d}\ (\boldsymbol{a}, \boldsymbol{b})$ represents the number of positions for which the condition is respected $\boldsymbol{a}(\boldsymbol{i}) \neq \boldsymbol{b}(\boldsymbol{i})$. In short, it is an estimate of how the two words differ from each other.
- **Hamming weight:** Given a word to, the weight of Hamming (HW) $\boldsymbol{d}\ (\boldsymbol{a}, \boldsymbol{0})$**,** represents the number of positions of where $a(i) \neq 0$.

3.1.1 Uniqueness

As explained above, the purpose of the PUFs is to generate a non-volatile key for greater security in device authentication, exploiting random phenomena, such as for example PUF process variations on silicon. It is important, however, that these keys are unique for each device as a sort of fingerprint and to verify it one can use the uniqueness understood as the degree of similarity between the responses of different PUFs to the same incoming challenge. To do this you can use the Hamming distance between the keys generated by two different devices (inter-HD) in order to have an estimate of how different the answers are when the same challenge is applied at the input. Taking into consideration an ideal random distribution the Hamming distance between two words generated by different chips should be close to 50%. From the Hamming distance it is possible to calculate the uniqueness of the response of a device through a simple formula. Considering the responses of k different devices $R_i(n)$ e $R_j(n)$ at n challenge it is possible to calculate the average Hamming distance between the two devices as follows:

$$HD_{inter}(\%) = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \frac{HD(R_i(n), R_j(n))}{n} * 100$$

For example, in Figure 19. We have two different devices that generate two to seven bits responses. It can be observed how, applied the same input challenge, the number of bits of the responses that differ is equal to 2 so, in this case, the Hamming distance is equal to (2/7) * 100 which is approximately equal to 28 % which is less than 50%.

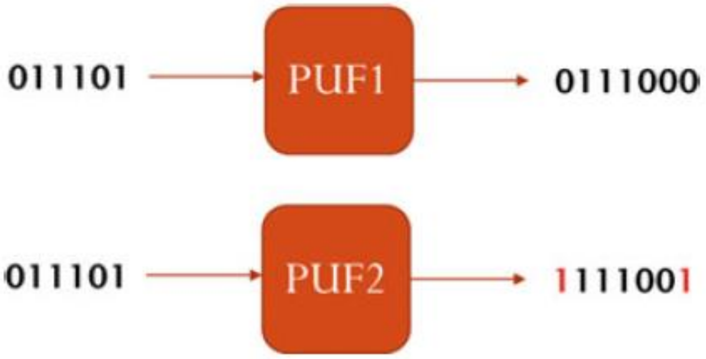


Figure 19. Example of the evaluation of the uniqueness of a PUF

3.1.2 Reliability

The generation of a key as a response to a challenge should ideally be independent of the working conditions in which it is generated. Due to the sensitivity of the behavior of many devices, from the temperature this aspect becomes critical in the design phase. Making a PUF stable to changes in operating conditions, such as variations in supply voltage or temperature, is one of the most important challenges to overcome during the design phase. This is because, for example, during the authentication phase the authenticator compares the PUF response of the device to be authenticated with those it previously stored in its database. But they refer to certain operating conditions and if the behavior of the PUF changes as the latter changes there would be errors during the authentication phase itself. It therefore becomes important to make the device insensitive to variations in temperature and voltage or at least to minimize the number of bits that flip (which change their logical state) as they vary so that possible errors can be corrected by means of correction circuits. error (ECC). To assess the reliability of a device, the Hamming distance (intra-HD) is usually used. Suppose we stimulate the device with the same challenge but vary the working conditions in which it is found, and suppose that this chip generates k answers at n bit R_i (n) then averaging the Hamming distance between the answers obtained by changing the working conditions and the one obtained under nominal conditions as follows:

$$HD_{intra}(\%) = \frac{1}{k}\sum_{i=1}^{k}\frac{HD(R_i(n), R'_j(n))}{n} * 100$$

The reliability of a PUF can be calculated as follows:

$$Affidabilit\acute{a} = 100\% - HD_{intra}$$

Ideally, we would like this value to be 100%, we really need to try to approach it as much as possible to the ideal value in order to correct the few errors with specific circuits.

For example, in Figure 20. we can analyze two seven-bit responses of a PUF to the same challenge but in different operating conditions. As can be seen, these two answers differ by only one bit and therefore the Hamming distance is equal to (1/7) * 100, therefore about 14% and consequently the reliability of the device is equal to 100% - 14% is equal to 86%.

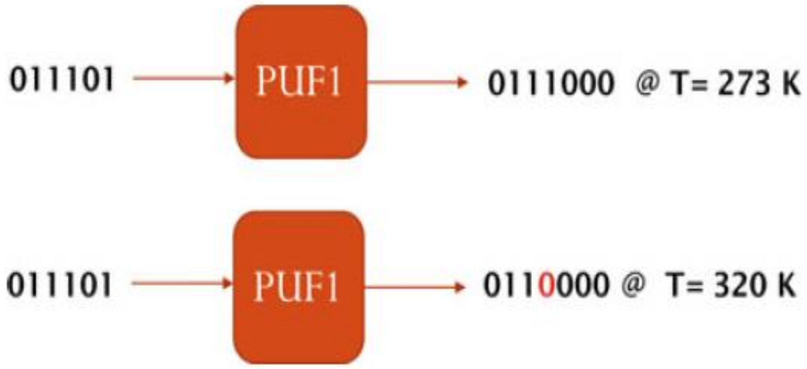


Figure 20. An example of evaluation of the reliability of a PUF

3.1.3 Uniformity

As mentioned in the introductory chapter, one of the main requirements for PUFs is randomness, that is, to make external attacks more difficult, it is important that the number of ones and zeros in the generated word is the same. To do this we can use the weight of Hamming which is a measure of the number of ones present in a word and we would ideally like it at 50%. To estimate the uniformity of the device it is enough to consider k responses of a PUF with the relative weight of Hamming r_i and do a media operation in the following way.

$$Uniformitá = \frac{1}{k}\sum_{i=1}^{k} r_i * 100\%$$

Obviously, we would also like in this case that the value was as close as possible to 50%.

3.1.4 Handling Strength

The purpose of the PUF is to replace authentication based on an identifier stored in memory to address the vulnerability of the same from invasive physical attacks. It becomes important therefore that the behavior of each PUF is unique, that is if a PUF is somehow replicated then it is important that the behaviors are different. Also, in this case the calculation to do is based on a sort of average of the Hamming distances of the responses of two different chips to the same challenge. The calculation is very similar to that performed for one of the previous metrics, except that in this case it is extended to all the possible challenge response pairs that can be generated by the device. So, considering the Hamming distance of the answers $R_i\ (l)$, $R_j\ (l)$ obtained with a challenge l it is possible to estimate the singularity of the behavior of a PUF, in the following way.

$$HD_{ave} = \frac{1}{CRP}\sum_{l=1}^{CRP}\frac{HD(R_i(l), R_j(l))}{n} * 100\%$$

In which by CRP we mean the total number of challenge response pairs that can be generated by the same.

Ideally the value should be 50% and this would indicate that the device is resistant to any invasive physical attacks.

3 Conclusions

In conclusion, we started by trying to get an overview of the physical unclonable functions and the possible application implications, starting from the assumption that with the development of the Internet of thing and of the growing number of interconnected devices, hardware and software security is covering an ever increasing aspect crucial. Physical unclonable functions are devices capable of exploiting intrinsic causal phenomena to generate unique keys for each device as a sort of fingerprint. The objective of the paper was to analyze the state of the art of PUF, also underlining some crucial aspects in the design phase, and how they are used to obtain a safe and low-cost authentication phase, through challenge mechanisms response, and to generate encryption keys. Various types of PUF were analyzed, from optical ones, which exploit the random scattering of light to generate the key, to those based on silicon, which exploit the process variations intrinsic to the manufacturing process to generate the key, and were analyzed some crucial aspects in the manufacturing phase. Finally, some metrics were presented to assess the quality of a PUF's design considering all the challenges that must be addressed during their design. In conclusion, PUFs have great commercial potential, but there are still many design challenges that must be overcome before reliable products can be obtained. What is expected is that in the future further technological developments, for example the technological scaling of silicon-based technologies, will favor the design of new PUFs.

4 Bibliography

[1] J. W. Lee, D. Lim, B. Gassend, G. E. Suh, M. Van Dijk, and S. Devadas, "A technique to build a secret key in integrated circuits for identification and authentication applications," *IEEE Symp. VLSI Circuits, Dig. Tech. Pap.*, no. CIRCUITS SYMP., pp. 176–179, 2004.

[2] B. Halak, M. Zwolinski, and M. S. Mispan, "Overview of PUF-Based hardware security solutions for the internet of things," *Midwest Symp. Circuits Syst.*, no. October, pp. 1–4, 2017.

[3] R. De Rose, F. Crupi, M. Lanuzza, and D. Albano, "A physical unclonable function based on a 2-transistor subthreshold voltage divider," *Int. J. Circuit Theory Appl.*, vol. 45, no. 2, pp. 260–273, 2017.

[4] U. Rührmair and D. E. Holcomb, "PUFs at a glance," *Proc. -Design, Autom. Test Eur. DATE*, 2014.

[5] C. Herder, M. D. Yu, F. Koushanfar, and S. Devadas, "Physical unclonable functions and applications: A tutorial," *Proc. IEEE*, vol. 102, no. 8, pp. 1126–1141, 2014.

[6] G. E. Suh and S. Devadas, "REF_6_Physical Unclonable Functions for Device Authentication," 2007.

[7] R. M. Intrinsic-id, *Physically Unclonable Functions : A Study on the State of the Art and Future Research Directions Physically Unclonable Functions : a Study on the State of the Art and Future Research Directions .*, no. December 2009. 2016.

[8] U. Rührmair, C. Hilgers, and S. Urban, "Optical PUFs Reloaded," *IACR Cryptol.*, 2013.

[9] S. Eiroa, I. Baturone, A. J. Acosta, and J. Davila, "Using Physical Unclonable Functions for Hardware Authentication: A Survey," *XXV Conf. Des. Circuits Integr. Syst.*, vol.

248858, pp. 204–209, 2010.

[10] S. Ghosh, “Spintronics and Security: Prospects, Vulnerabilities, Attack Models, and Preventions,” *Proc. IEEE*, vol. 104, no. 10, pp. 1864–1893, 2016.

[11] I. Verbauwhede and R. Maes, *Physically unclonable functions*. 2011.

[12] K. Lofstrom, W. R. Daasch, and D. Taylor, “IC identification circuit using device mismatch,” *Dig. Tech. Pap. - IEEE Int. Solid-State Circuits Conf.*, vol. 46, no. 8, pp. 372–373, 2000.

[13] M. Bogers, “Business Model Innovation II Module 18 in “ Innovation & Knowledge “ Introduction to Business Model Innovation I & II,” pp. 9–14, 2016.

[14] M. Kalyanaraman and M. Orshansky, “Novel strong PUF based on nonlinearity of MOSFET subthreshold operation,” *Proc. 2013 IEEE Int. Symp. Hardware-Oriented Secur. Trust. HOST 2013*, pp. 13–18, 2013.

[15] J. Guajardo, S. S. Kumar, G. Schrijen, and P. Tuyls, “FPGA Intrinsic PUFs and Their Use for IP Protection.”

[16] J. Lee, D.-W. Jee, and D. Jeon, “Power-up control techniques for reliable SRAM PUF,” *IEICE Electron. Express*, vol. 16, no. 13, pp. 20190296–20190296, 2019.

[17] A. Shamsoshoara, A. Korenda, F. Afghah, and S. Zeadally, “A Survey on Hardware-based Security Mechanisms for Internet of Things,” pp. 1–33, 2019.