UNIVERSITÀ DELLA CALABRIA

UNIVERSITA' DELLA CALABRIA

Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica

Dottorato di Ricerca in

Information and Communication Technologies

CICLO

XXXV

DESIGN OF PHYSICALLY UNCLONABLE FUNCTIONS IN CMOS AND EMERGING TECHNOLOGIES FOR HARDWARE SECURITY APPLICATIONS

Settore Scientifico Disciplinare ING-INF/01

Coordinatore: Ch.mo Prof. Giancarlo Fortino

Firma _ Firma oscurata in base alle linee guida del Garante della privacy

Supervisore/Tutor: Ch.mo Prof. Felice Crupi

Firma___ Firma oscurata in base alle linee guida del Garante della privacy

Dottorando: Dott. Massimo Vatalaro

Firma _ Firma oscurata in base alle linee guida del Garante della privacy

Preface

It has been three years since I started my PhD. There are no words to summarize this important period of my life. In these three years I met many wonderful people who allowed me growing continuously.

First of all, I would thank my relator Prof. Felice Crupi for guiding me in the best way in these three years. Thank you for teaching me so much and for always encouraging me on this journey.

Special thanks to Dr. Raffaele De Rose whose help and support allows me growing professionally and personally. Your inspiring discussion and guidance are greatly appreciated.

Thanks to Prof. Marco Lanuzza for all the support you have given me in these three years.

I would thank my friends and colleagues Vincenzo Maccaronio, Fanny Spagnolo and Stefano Laureti for their help and for making even the most difficult moments light. Many thanks to my friend Cristian Sestito for supporting me every day and for all the suggestions and help in these three years.

I also want to express my greatest gratitude to my parents and my sister. Thank you for being the strongest support during this important journey. I would not succeed on pursuing this degree without your encouragement.

Finally, special thanks to my life partner Wiky. My life would not be the same without your daily love and support. Thank you for pushing me to always give my best.

Abstract

The advent of the IoT scenario heavily pushed the demand of preserving the information down to the chip level due to the increasing demand of interconnected devices. Novel algorithms and hardware architectures are developed every year with the aim of making these systems more and more secure. However, IoT devices operate with constrained area, energy and budget thus making the hardware implementation of these architectures not always feasible. Moreover, these algorithms require truly random key for guarantying a certain security degree. Typically, these secret keys are generated off chip and stored in a non-volatile manner. Unfortunately, this approach requires additional costs and suffers from reverse engineering attacks. Physically unclonable functions (PUFs) are emerging cryptographic primitives which exploit random phenomena, such as random process variations in CMOS manufacturing processes, for generating a unique, repeatable, random, and secure keys in a volatile manner, like a digital fingerprint. PUFs represent a secure and low-cost solution for implementing lightweight cryptographic algorithms. Ideally PUF data should be unique and repeatable even under noisy or different environmental conditions. Unfortunately, guarantying a proper stability is still challenging, especially under PVT variations, thus requiring stability enhancement techniques which overtake the PUF itself in terms of required area and energy. Nowadays, different PUF solutions have been proposed with the aim of achieving ever more stable responses while keeping the area overhead low.

This thesis presents a novel class of static monostable PUFs based on a voltage divider between two nominally identical sub-circuits. The fully static behavior along with the use of nominally identical sub-circuits ensure that the correct output is always delivered even when on-chip noise occasionally flips the bit, and that randomness is always guaranteed regardless of the PVT conditions. Measurement results in 180-nm CMOS technology demonstrates the effectiveness of the proposed solution with a native instability (BER) of only 0.61% (0.13%) along with a low sensitivity to both temperature and voltage variations. However, these results were achieved at the cost of more area-hungry design (i.e., $7{,}222F^2$) compared to other relevant works. The proposed solution was also implemented with emerging paper based MoS_2 nFETs by exploiting a LUT-based Verilog-A model, calibrated with experimental I_D vs V_{DS} at different V_{GS} curves, whose variability was extracted from different I_D vs V_{GS} curves of 27 devices from the same manufacturing lot. Simulations results demonstrate that these devices can potentially used as building block for next generation electronics targeting hardware security applications. Finally, this thesis also provides an application scenario, in which the proposed PUF solution is employed as TRNG module for implementing a smart tag targeting anti-counterfeiting applications.

Sommario

Il continuo sviluppo dell'IoT e il conseguente incremento del numero di dispositivi connessi tra di loro aumentano drasticamente la richiesta di garantire un certo grado di sicurezza anche a livello hardware. Con il passare degli anni, sempre più task vengono svolti digitalmente semplificando la vita di tutti i giorni. Tuttavia, questi benefici rappresentano anche punti di vulnerabilità che un utente male intenzionato può sfruttare per impossessarsi di informazioni sensibili. Al giorno d'oggi, diversi algoritmi di crittografia sono stati sviluppati insieme ad architetture hardware efficienti per garantire all'utente un certo livello di sicurezza. Tuttavia, non è sempre possibile implementare questi algoritmi nei dispositivi IoT a causa dei vincoli stringenti in termini di area, costi e consumo di energia con cui spesso si trovano ad operare. Inoltre, bisogna considerare che questi algoritmi richiedono chiavi sicure e realmente random. Quest'ultime sono tipicamente generate off-chip e memorizzate in maniera non volatile così che possano essere prelevate quando richiesto. Tale approccio è molto costoso e, allo stesso tempo, rappresenta un punto di vulnerabilità richiedendo l'utilizzo di circuiti di protezione, che però necessitano di essere alimentati anche quando il chip è spento, aumentando così il costo energetico complessivo del chip durante il suo ciclo di vita. Le funzioni fisicamente non clonabili (PUF) rappresentano primitive di crittografia emergenti che sfruttano fenomeni casuali come, per esempio, le variazioni di processo durante la fabbricazione di dispositivi CMOS per generare chiavi sicure, uniche, casuali e riproducibili in maniera volatile. L'ultimo punto, in particolare, si riferisce al fatto che queste chiavi sono funzione del circuito che le genera e non vengono, quindi, memorizzate da nessuna parte. In questo modo, esse operano solamente quando vengono richieste eliminando così la necessità di particolari circuiti di protezione. Per un corretto funzionamento, le PUF devono garantire chiavi uniche e ripetibili anche in condizioni diverse da quelle nominali e/o rumorose. Nonostante il grande interesse scientifico, garantire un'appropriata stabilità della risposta delle PUF è ancora una sfida, che spesso comporta la necessità di implementare tecniche di correzione esterne che possono sovrastare la PUF stessa sia in termini di area che di energia richiesta.

Questa tesi propone una nuova classe di PUF statiche e monostabili che sfrutta un partitore di tensione tra due sottocircuiti identici. Il comportamento statico assicura che la corretta uscita venga sempre garantita, anche quando il rumore cambia occasionalmente il bit di uscita, mentre l'utilizzo di due sottocircuiti nominalmente identici garantisce un'adeguata casualità della risposta a prescindere dalle condizioni di lavoro. I risultati di misura ottenuti in tecnologia CMOS a 180-nm confermano l'efficacia della soluzione proposta, mostrando una instabilità (BER) nativa di 0.61% (0.13%) ed un altrettanto bassa sensibilità alle variazioni della tensione di alimentazione e della temperatura. Tuttavia, questi risultati sono stati ottenuti a costo di una maggiore area ($7{,}222F^2$) rispetto ad altri lavori. Questa soluzione è stata anche investigata utilizzando dispositivi MoS_2 emergenti fabbricati con l'impiego della carta come substrato. In particolare, è stato sfruttato un modello LUT-based calibrato con dati sperimentali in cui le curve I_D-V_{DS} a differenti V_{GS} sono state sfruttate per modellizzare il comportamento elettrico dei dispositivi, mentre le curve I_D-V_{GS} di 27 dispositivi provenienti dallo stesso lotto di fabbricazione sono state sfruttate per estrarre informazioni statistiche sulla variabilità degli stessi dispositivi. Infine, questa tesi affronta anche lo scenario applicativo dell'anticontraffazione, in cui la PUF proposta è stata inserita come elemento base di uno smart tag.

List of Abbreviations

ACF	Autocorrelation Function.
AES	Advanced Encryption Standard.
APUF	Arbiter PUF.
ASIC	Application Specific Integrated Circuit.
BC	Bit Commitment.
BCH	Bose-Chauduri-Hocquenhem.
BD	Breakdown.
BEOL	Back-End-Of-Line.
BER	Bit Error Rate.
CM	Current Mirror.
CMOS	Complementary Metal-Oxide-Semiconductor.
CMP	Chemical mechanical planarization.
CRP	Challenge-Response Pair.
CT	Challenge Trigger.
CTAT	Complementary To Absolute Temperature.
CTC	Cycles To Collapse.
CVD	Chemical Vapor Deposition.
DEA	Data Encryption Algorithm.
DES	Data Encryption Standard.
DFF	Delay Flip Flop.
DIBL	Drain-Induced Barrier Lowering.
DPA	Differential Power Attack.
ECC	Error Correction Code.
ECC	Elliptic Curve Cryptography.
ECDSA	Elliptic Curve Digital Signature Algorithm.
EE SRAM	Enhancement-Enhancement Static Random Access Memory.

EGFET	Extended Gate Field-Effect Transistor.
EM	ElectroMagnetic
FEOL	Front-End-Of-Line.
FET	Field-Effect Transistor.
FPGA	Field-Programmable Gate Array.
FRAM	Flash-RAM.
FSM	Finite State Machine.
GK	Golden Key.
HCI	Hot-Carrier Injection.
HD	Hamming Distance.
HW	Hamming Weight.
IC	Integrated Circuit.
ID	Identity Document
IoT	Internet of Things.
IP	Intellectual Property.
KE	Key Exchange.
KER	Key Error Rate.
LUT	Look-Up Table.
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor.
MTBF	Mean Time Before Failure.
MTJ	Magnetic Tunnel Junction.
MUX	Multiplexer.
MVT	Medium Threshold Voltage.
NAND	Not-and.
NBTI	Negative Biased-Temperature Instability.
NFC	Near Field Communication.
NIST	National Institute of Standard and Technology.
NOR	Not-or.

NRAM	Nano-RAM.
NVM	Non-Volatile Memory.
OT	Oblivious Transfer.
OTP	One Time Programmable Memory.
PBTI	Positive Biased-Temperature Instability.
PDN	Pull-Down Network.
PQ	Physical Quantity.
PRNG	Pseudo Random Number Generator.
PVC	Process-to-Voltage Converter.
PVT	Process-Voltage-Temperature.
PTC	Process-to-Time Converter.
PUF	Physically Unclonable Functions.
PUN	Pull-Up Network.
RCCM	Regulated Cascode Current Mirror.
RE	Reverse Engineering.
RFID	Radio Frequency Identification.
RO	Ring Oscillator.
ReRAM	Resistive Random Access Memory.
RSA	Rivest-Shamir-Adleman.
RTL	Register Transfer Level.
SA	Sense Amplifier.
SBD	Soft Breakdown.
SBOX	Substitution Box.
SCA	Sub-threshold Current Array.
SHA	Secure Hash Algorithm.
SMV	Spatial Majority Voting.
SNR	Signal-to-Noise Ratio.
SPA	Simple Power Attack.

SRAM	Static Random Access Memory.
TMV	Temporal Majority Voting.
TRNG	True Random Number Generator.
XNOR	Exclusive-Nor.
XOR	Exclusive-or.

OUTLINE

List of figures

I_{2T}) (c) simulated across voltages at T = 25°C from 250 samples and (d) measured across voltages at T = 25°C from 20 dice.

3.7 *V_X voltage of the 2T-core normalized to V_{DD} (a) simulated across temperatures at V_{DD}= 1.8 V from 250 samples and (b) measured across temperatures at V_{DD}= 1.8 V from 20 samples. Absorbed current from the 2T core (i.e., I_2T) (c) simulated across temperatures at V_{DD}= 1.8 V from 250 samples and (d) measured across temperatures at V_{DD}= 1.8 V from 20 dice.*

3.8 *Statistical distributions of (a) the DIBL coefficient difference ($\Delta\lambda_{D1,2}$), (b) the V_{TH0} difference ($\Delta V_{TH1,2}$) and (c) the V_{TH} temperature coefficient difference ($\Delta k_{T1,2}$) for a PMOS MVT device with W = 0.22 µm and L = 0.25 µm (from 5k-run Monte Carlo simulations under local variations in the considered 180-nm technology). Statistical correlation between (d) DIBL coefficient mismatch ($\Delta\lambda_{D1,2}$) and threshold voltage mismatch ($\Delta V_{TH01,2}$) and (e) V_{TH} temperature coefficient mismatch ($\Delta k_{T1,2}$) and threshold voltage mismatch ($\Delta V_{TH01,2}$) for two PMOS MVT devices with W = 0.22 µm and L = 0.25 µm (from 5k-run Monte Carlo simulations under local variations in the considered 180-nm technology).*

3.9 *(a) Schematic and (b) Layout of the PUF bitcell based on the 2T sub-threshold voltage divider.*

3.10 *Simulation results (5k-run Monte Carlo at V_{DD} = 1.8 V and T = 25°C) of 2T-core PUF bitcell in 180-nm CMOS at TT corner: (a) statistical distribution of the voltage V_X of the bitcell core, (b) nominal input-output characteristics of the inverter, and (c) statistical distribution of the voltage V_{OUT} of the inverter.*

3.11 *Effect of voltage and temperature variations on the static parameters of the output inverter at different process corners. Effect of (a) voltage with T= 25 °C, and (b) temperature with V_{DD} = 1.8 V variations on the difference between minimum high- and maximum low-input voltages (V_{IH}-V_{IL}). Effect of (c) voltage with T= 25 °C and (d) temperature with V_{DD} = 1.8 V variations on the input logic-threshold (V_M).*

3.12 *Mean value and standard deviation of (a) V_{TH0}, (b) DIBL coefficient (λ_D), and (c) V_{TH} temperature coefficient (k_T) from 5k-run Monte Carlo simulations at different process corners.*

3.13 *Statistical distribution of the V_X voltage of the 2T-core from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T = 25°C) at (a) TT corner, (b) FF corner, and (c) SS corner.*

3.14 *Percentage of simulated unstable bits for the 2T-based solution at TT corner under (a) V_{DD} variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V.*

3.15 *Percentage of total unstable bits for the 2T-based solution across different process corners under (a) voltage variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V.*

3.16 *Simulated absorbed current by the 2T-core across different process corners from 5k-run Monte Carlo simulations under (a) voltage variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V. Simulated absorbed currend across different process corners by the bitcell under (c) voltage variations at T= 25°C and (d) temperature variations at V_{DD}= 1.8 V.*

3.17 *(a) Schematic of the bitcell based on the 4T voltage divider along with (b) the operative principle.*

3.18 *V_X variability as function of (a) $L_{3,4}$ with $W_{3,4}$ = 0.25 µm, (b) $W_{3,4}$ with $L_{3,4}$ = 0.25µm, and (c) both $L_{3,4}$ and $W_{3,4}$ at nominal conditions of V_{DD} = 1.8 V and T = 25 °C from 1k-run Monte Carlo simulations.*

3.19 *V_X variability as function of the M3-M4 channel widths with L = 0.5 µm at GK conditions (i.e., V_{DD} = 1.8 V and T = 25 °C) from 1k-run Monte Carlo simulations.*

3.20 *Statistical distribution of (a) DIBL coefficient (λ_D), (b) Threshold voltage (V_{TH0}), and (c) threshold voltage temperature coefficient (k_T) from 5k-run Monte Carlo simulations for a PMOS MVT device with nominal sizing (i.e., L = 0.5 µm and W = 1.5 µm).*

3.21 *(a)V_X Voltage as function of the M1-M2 mismatch at GK conditions (i.e., V_{DD} = 1.8 V and T = 25°C) and (b) gain (i.e., $(1+\lambda_{D3,4})/(2\lambda_{D1,2}\lambda_{D3,4})$) distribution from 5k-run Monte Carlo simulations.*

3.22 *Percentage of samples in which M1-M2 mismatch overtakes the M3-M4 mismatch as function of $W_{3,4}$ sizing with $L_{3,4}$ = 0.5 µm at GK conditions (i.e., V_{DD}= 1.8 V and T = 25 °C) from 5k-run Monte Carlo simulations.*

3.23 *Voltage drops across M1-M4 transistors in the 4T voltage divider normalized to the V_{DD} for strong M1-M2 mismatch and weak M1-M2 mismatch. (a) Strong logic '0', (b) weak logic '0', (c) strong logic '1', and (d) weak logic '1'.*

3.24 *V_X voltage of the 4T-core normalized to V_{DD} (a) simulated across voltages at T = 25°C from 250 samples and (b) measured across voltages at T = 25°C from 20 samples. Absorbed current from the 4T core (i.e., I_{4T}) (c) simulated across voltages at T = 25°C from 250 samples and (d) measured across voltages at T = 25°C from 20 dice.*

3.25 *V_X voltage of the 4T-core normalized to V_{DD} (a) simulated across temperatures at V_{DD}= 1.8 V from 250 samples and (b) measured across temperatures at V_{DD}= 1.8 V from 20 samples. Absorbed current from the 4T core (i.e., I_4T) (c) simulated across temperatures at V_{DD}= 1.8 V from 250 samples and (d) measured across temperatures at V_{DD}= 1.8 V from 20 dice.*

3.26 *(a) Schematic and (b) Layout of the PUF bitcell based on the 4T sub-threshold voltage divider.*

3.27 *Simulation results (5k-run Monte Carlo at V_{DD} = 1.8 V and 25 ◦C) of 2T-core versus 4T-core PUF bitcell in 180-nm CMOS: (a) statistical distribution of the voltage V_X of the bitcell core, (b) nominal input–output characteristics of the inverter, and (c) statistical distribution of the voltage V_{OUT} of the inverter.*

3.28 *Mean value of (a) V_{TH0}, (b) DIBL coefficient (λ_D), and (c) V_{TH} temperature coefficient (k_T) of M1-M2 and M3-M4 transistors from 5k-run Monte Carlo simulations at different process corners.*

3.29 *Statistical distribution of the V_X voltage of the 4T-core from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T = 25°C) at (a) TT corner, (b) FF corner, and (c) SS corner.*

3.30 *Mean values of $|V_{SD1} - V_{SD2}|$ and $|V_X - V_{DD}/2|$ of the 4T-core at different process corners from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C).*

3.31 *Percentage of simulated unstable bits for the 4T-based solution from 5k-run Monte Carlo simulations at TT corner under (a) V_{DD} variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V.*

3.32 *Statistical distribution of the voltage V_X of the 4T bitcell core at different VT corners from 5k-run Monte Carlo simulations: (a) V_{DD}= 1.8 V and T = 0 °C, (b) V_{DD}= 1.8 V and T = 100 °C, (c) V_{DD}= 0.4 V and T = 0 °C, and (d) V_{DD}= 0.4 V and T =100 °C.*

3.33 *Percentage of total unstable bits for the 4T-based solution across different process corners under (a) voltage variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V.*

3.34 *Simulated absorbed current by the 4T-core across different process corners from 5k-run Monte Carlo simulations under (a) voltage variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V. Simulated absorbed currend across different process corners by the bitcell under (c) voltage variations at T= 25°C and (d) temperature variations at V_{DD}= 1.8 V.*

3.35 *Architecture of the PUF array.*

3.36 *(a) Photograph of the packaged test chip and layouts of (b) 8 × 32 PUF array, and (c) PUF bitcell area including pass transistor.*

3.37 *Measurements of the 8 × 32 PUF array at GK conditions (V_{DD}= 1.8 V and T= 25 °C) across seven dice: (a) percentage of bit '0', bit '1', and unstable bits; (b) logical speckle diagram; (c) percentage of unstable bits versus number of evaluations, and (d) unstable bit mask at 500 evaluations.*

3.38 *Percentage of unstable bits (averaged over seven dice) under (a) V_{DD} variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V.*

3.39 *Example of M1 and M2 threshold voltages trend under temperature variations when (a) $k_{T,1} = k_{T,2}$, (b) $k_{T,1} < k_{T,2}$, and (c) $k_{T,1} > k_{T,2}$.*

3.40 *BER under (a) V_{DD} variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V. Data are averaged over seven dice considering 32-bit PUF words.*

3.41 *(a) Normalized inter-PUF and intra-PUF HD at GK conditions (i.e., due only to on-chip noise), (b) normalized intra-PUF HD under voltage and temperature variations, (c) normalized number of bit "1" at GK conditions. Data are evaluated across seven dice considering 32-bit PUF words, and (d) spatial autocorrelation function (ACF).*

3.42 *Measured supply current (I_{DD}) per bitcell versus V_{DD} at T= 25 °C for (a) six of seven dice, and (b) averaged over seven dice with relative static power.*

3.43 *Schematic of (a) high-gain 4T inverter design, and (b) low-area 4T inverter design.*

3.44 *Effect of voltage and temperature variations on the static parameters of the output inverter for both high-gain design and low-area design. Effect of (a) voltage with T = 25 °C and (b) temperature with V_{DD} = 1.8 V variations on the difference between minimum high- and maximum low-input voltages (V_{IH}-V_{IL}). Effect of (c) voltage with T = 25 °C and (d) temperature with V_{DD} = 1.8 V variations on the input logic-threshold (V_M).*

3.45 *Statistical distribution of the amplitude of the unstable region (i.e., V_{IH} - V_{IL}) and the input logic logic threshold (i.e., V_M) at (a) and (b) V_{DD}= 1.8 V and T= 25 °C, and (c) and (d) V_{DD}= 0.4 V and T= 25 °C, respectively, from 5k-run Monte Carlo simulations at TT corner.*

3.46 *Short-circuit current (i.e., when $V_X = V_M$) of the inverter under (a) V_{DD} variations at T= 25 °C and (b) Temperature variations at V_{DD}= 1.8 V.*

3.47 *(a) Bitcell design concept along with schematic and layout (b) without using the body effect and (c) using the body effect.*

3.48 *Comparison at GK conditions (i.e., V_{DD} = 1.8 V and T = 25°C) between the two solutions in terms of (a)V_X voltage and (b) $V_{SD1} - V_{SD2}$ as function of the M1-M2 mismatch, and (c) gain distribution. Data came from 5k-run Monte Carlo simulations.*

3.49 *Effect of voltage and temperature variations on the static parameters of the output inverter at different process corners. Effect of (a) voltage with T= 25 °C, and (b) temperature with V_{DD} = 1.8 V variations on the difference between minimum high- and maximum low-input voltages (V_{IH}-V_{IL}). Effect of (c) voltage with T= 25 °C and (d) temperature with V_{DD} = 1.8 V variations on the input logic-threshold (V_M).*

3.50 *Percentage of unstable bits ('noisy'+ 'flipped') at different process corners across voltages (T = 25 °C) and temperatures (V_{DD}= 1.8 V). (a) and (c) refer to the bitcell of Fig. 3.47(b), (b) and (d) refer to the bitcell of Fig. 3.47(c).*

3.51 *Percentage of unstable bits averaged over process corners across (a) voltages and (b) temperatures for the two circuits of Fig. 3.47(b)-(c), (c) summary comparison in terms of unstable bits under different PVT conditions.*

3.52 *Flipping probability as function of both input−output characteristic of the inverter of Fig.3.43(b) and V_X distribution.*

3.53 *Summary comparison between the two circuits of Fig. 3.47(b)-(c) in terms of bit error rate (BER) under different PVT conditions. Data refers to 32-bit PUF words within an 8×32 bitcell array.*

3.54 *Summary comparison between the two circuits of Fig. 3.47(b)-(c) in terms of static power consumption per bitcell under different PVT conditions. BER data refers to an 8×32 bitcell array with power gating.*

3.55 *(a) Bitcell design concept along with schematic and layout (b) 6T-core based bitcell and (c) 8T-core based bitcell.*

3.56 *Comparison at GK conditions (i.e., V_{DD} = 1.8 V and T = 25°C) between 6T-core and 8T-core circuits in terms of (a)V_X voltage as function of the M1-M2 mismatch, and (b) gain distribution. Data came from 5k-run Monte Carlo simulations.*

3.57 *Voltage drops across the transistors in the (a) and (c) 6T-core voltage divider and (b) and (d) 8T-core voltage divider, normalized to the V_{DD} for weak M1-M2 mismatch. (a) and (b) weak logic '0', (c) and (d) weak logic '1'.*

List of Tables

Chapter 1
Introduction

With the advent of the Internet of Things (IoT) scenario the number of interconnected devices is increasing strongly. Nowadays, the IoT paradigm includes more than 20 billion connected devices which are expected to be used in a wide variety of applications such as personal health monitoring, smart home, smart cars, environmental monitoring systems and critical infrastructure, and so on [1]. In addition, increasingly critical tasks are entrusted to portable devices such as shopping, bank transaction and business. This continuous exchange of information is enabled by the network connectivity and can be controlled remotely thus highlighting that many challenges need to be faced along with these benefits. While improving our life a huge amount of information is stored and transferred in the IoT network thus leading to some side effect. The data includes private and critical information whose leaking will lead to threats to system security and user privacy. As a results, security must be guaranteed in the communication channels as well as secure authentication protocols need to be adopted for ensuring that a user is who he says he is.
Information and network security refer to the protection of the information that is stored, transmitted, and processed in a networked system [2]. There are three key concepts which embody the fundamental security objectives:

- ***Confidentiality*** preserves the information from an unauthorized access and disclosure.
- ***Integrity*** provides protection against improper information modification.
- ***Availability*** ensures reliable and timely access to the information.

These three terms refer to the CIA triad and represent security objectives for the information security. Someone in the security field added additional concepts such as

- ***Authenticity*** controls that the source of the information is a trusted source and verifies that users are who they say they are.
- ***Accountability*** records entity activities for permitting later forensic analysis.

These security objectives can be ensured through some security mechanisms such as cryptographic algorithm, data integrity, digital signature, authentication exchange, traffic padding, routing control, and so on. Many of these mechanisms are implemented by well-established cryptographic algorithms, which can be divided into three categories: keyless (i.e., without any keys during cryptographic transformations), single-key (i.e., the transformed data is function of the input data and a secret key) and two-key (i.e., when transforming input data two different but correlated keys are used, private key and public key). The first class is mainly used for turning a variable quantity of text into a fixed-length and apparently random value known as hash value, hash code, or digest. Since the variations of one bit will result, with high probability, in a different hash value this cryptographic function presents three interesting properties for which (*i*) it is infeasible to find a data that maps a pre-specified hash result and (*ii*) it is infeasible to find two data that map the same hash value and (*iii*) this algorithm can be broken only with brute-force attacks (i.e., a malicious user tries every possible key combination for breaking the algorithm). Indeed, these functions are often used for determining changes in the data thus allowing of verifying the data integrity. The most widely used hash function is the secure hash algorithm (SHA). During the years different versions have been developed for facing the technological

development with the relative increase of the effectiveness of the brute-force attacks. These properties make this class useful for applications such as message authentication, digital signature, and pseudorandom number generation (PRNG). The second class implies the use of a secret key belonging to a single user. Typically, this key is shared between two or more parities for implementing the encryption algorithm well known as symmetric encryption algorithms (or symmetric cipher). Indeed, these encryption algorithms receive input data and secret key and implement an intelligible transformation on the data. The decryption algorithms will recover the original data from the transformed one and the shared key. Symmetric encryption can operate on data as sequence of blocks (block ciphers) or as a sequence of bits (stream ciphers). Until the 2001 the most used symmetric cipher was the data encryption standard (DES) where the data encryption algorithm (DEA) performs the data transformation from the plaintext to the ciphertext by operating with 64-bit data blocks and a 64-bit private key. The algorithm performs permutation and substitution functions in different rounds along with circular shifts for translating the input data in an intelligible way. Since the 2001 the DES was replaced by the advanced encryption standard (AES) by the national institute of standards and technology (NIST) for a wide range of applications. The algorithm is referred to AES-128, AES-192 or AES-256 according to the key length and performs permutation and substitution functions operating with 16 bytes (4x4 matrix) for the intelligible transformation. Finally, the third class involves the use of two related keys: private and public. The first one is known only by a single user, whereas the latter is made available to the other users. These encryption schemes are well known as asymmetric encryption algorithms (or asymmetric cipher) and operates in two ways:

- Sender translates the input data in an intelligible way by using the private key. Subsequently, the receiver recovers the input data by performing the decryption algorithm involving the public key.
- Sender translates the input data in intelligible way by using the public key. Later, the receiver performs the decryption algorithm with the private key.

The most widely used algorithm is the Rivest-Shamir-Adleman (RSA) scheme which is a general-purpose approach to public-key encryption. This type of algorithms is mainly used for applications like digital signature, key exchange, and user authentication.
These algorithms need to be physically implemented reducing the hardware vulnerabilities.

1.1 Hardware Security

Designing secure systems while meeting at the same time IoT paradigm constraints is not an easy task. Indeed, with the progress in nanotechnologies, breaking cryptographic algorithms becomes faster thus requiring more and more complex security protocols which often cannot be implemented in an IoT device with limited budget in terms of battery energy, manufacturing cost and area occupation. This increases the complexity during the design phase since adding countermeasures worsens the overall cost and required energy budget [3], [4]. IoT network is composed by devices which operate using battery or scavenged energy thus indicating that particular attention must be paid for improving energy efficiency and reducing the required power

consumption for each task. Anyway, protecting privacy, authenticating data or sources of information, providing resistance to physical manipulation requires of adding security, cryptographic capability and other countermeasures to the IC design and it is not a simple task. These circuits must be energy efficient and compact but at the same time they must guarantee protection against physical attacks and avoid any leakage of the sensitive information during implementation of a security algorithm. For better understanding the complexity of adding security in IC design we can examinate two possible scenario. We can assume a network between two characters such as Alice and Bob with a malicious user which performs attacks only on the communication channel between the two parties (black box attacker model). In this scenario Alice intends to share confidential information with Bob over an insecure channel where eavesdropping can be performed by a malicious user for extracting the plaintext/ciphertext pairs. Here, Eve tries to guess the secret key with which the translation has been performed. In this case security strength is strongly related to the computational complexity of the underlying cryptographic algorithms. In this scenario if Eve succeeds in guessing the secret key faster than using brute-force attacks (i.e., trying all possible combinations) the algorithm can be considered broken. Nowadays the scaling of nanotechnology and the increase of the computational power leads to use more and more longer secret key for increasing the required time for an attacker to try all possible secret key combinations. Indeed, for long term security the suggested key lengths are 256-bit for secret key size, 512-bit for hash output size, 15,360 for the RSA modulus size and 512 for bit elliptic curves. Today the IoT paradigm includes billions of devices distributed everywhere allowing us talking about smart home, smart cars, wearable sensors and so on. This strong development of the distributed electronics has some side effects since more electronics also refers to more vulnerabilities. In this scenario we can assume that an attacker has access to both the communication channel and the devices (gray box attacker model) thus complicating more the design phase of modern ICs. Starting from the design phase to the system integration there are multiple points within this supply chain which could represent vulnerability points for an attacker. Some of these hardware-based threats are, for example [4]: hardware trojans, IP piracy and IC overbuilding, reverse engineering (RE), side-channel attacks, and counterfeiting. Hardware Trojans refer to the malicious circuit modifications [4]. In particular, the They may control, monitor, disable or modify the contents and communications of the underlying circuits by adding, for example, Trojans into the designs by manipulating the lithographic masks. In this case these Trojans assume form of addition, deletion or modification of gates and the detection is difficult for different reason:

- The opaqueness of the IC internal hurdles limits the detection of modified components.
- The technological scaling with the respective nondeterministic behavior makes more difficult to distinguish between process variations and Trojans hard.
- There is a large space in the IC for placing possible Trojans.

There are two possible approaches as countermeasures to these threats: invasive and non-invasive. The former makes the can potentially make the devices under test unusable later and requires high cost and precision. Indeed, it can be done only by big silicon companies. On the other hand, the non-invasive methods consist of testing circuits with pre-established patterns while controlling the respective output as well as the side-channel effects (i.e., delay, leakage, power, thermal profiling, etc.) [4]. IP piracy and IC overbuilding refer to claiming and overbuilding an IP and IC respectively. As countermeasure five methods have been developed:

- *Obfuscation*, which refers to hide the correct functionality of the IC by adding additional gates into it. There are several types of obfuscation, some of them include the insertion of XOR/XNOR gates and memory elements, in which the obfuscated design will work correct when applying the correct values to these blocks, while other types of obfuscation include extra states in a FSM (for example unused, invalid and black-hole states).
- *Watermarking*, which consists of including a designer's signature in the design artifact such as black-hole states in the finite-state machine (FSM), some secret constraint during physical and logic synthesis. The designer can later reveal the watermark and claim ownership of an IC/IP. A watermark should be transparent to the circuit functionality and extremely difficult to remove, it should be a conclusive proof of ownership and it should be also applicable to all design.
- *Fingerprinting*, which helps to avoid piracy by embedding the signature of the buyer on the IC (for example the public key). This solution can be implemented along with the watermark so that when challenged the designer can reveal the watermark and the signature for claiming the ownership and revealing an eventual source of piracy respectively. One typical approach consists of using power, thermal or timing fingerprint of an IC. Recently, another approach under research consists of using emerging device such as physical unclonable functions (PUF) for exploiting random physical phenomena as static entropy source for generating a volatile chip ID.
- *Metering*, which refers to a set of tools, methodologies and protocols for tracking the IC. There is passive and active metering. The first one used part of an IC's functionality for metering while in the active metering some parts of the IC's functionality can be only accessed by the designer.
- *Split Manufacturing*, which consists of splitting the layout in front-end-of-line (FEOL) and back-end-of-line (BEOL). These two manufacturing processes are fabricated separately in different foundry and then aligned and integrated with electrical, mechanical and optical techniques. The FEOL includes the layout at transistor level and at lower levels of metal (i.e., $\leq$ M4) while the BEOL refers to the layout at higher levels of metal (i.e., $>$ M4). An attacker cannot guess the connections associated to the BEOL by knowing the FEOL layers.

Reverse engineering consists of extracting sensitive information such as technology, the gate-level netlist or the functionality of the IC with the aim of fully reversing engineer a design to the desired abstraction level for stealing the IP or copying the IC [4]. Some of the principal countermeasures are:

- *Obfuscation*, which is similar to that described above and consists of including additional gates and memory elements for hiding the original design and functionality.
- *Camouflaging,* which includes techniques for masking the IC design at layout-level. In particular, this class of countermeasures allows hampering the image-processing-based extraction of the gate-level netlist. Indeed, NAND and NOR gates can look like the same logic gate at the layout-level despite their different functionality. Another approach consists of filling the unused space with filler cells such as programmable standard cells or dummy contacts.

Side-channel attacks represent a crucial problem since they exploit physical quantities during the IC operations for extracting secret information such as the private key [4]. During the years these

attacks demonstrated of being powerful and able to break most existing important cryptographic algorithms. Timing, power consumption, electromagnetic (EM) emanations, photonic emission and acoustic noise of the system could be correlated to the processed data when implementing some cryptographic algorithms in hardware. This correlation can be exploited for extracting sensitive and crucial information. In particular, in a timing attack a malicious user observes differences in execution time when processing the private key or sensitive data. Indeed, during the hardware implementation of the substitution boxes (Sboxes) if the data in a cache depend on the private key an attacker can exploit timing differences for extracting the key. Another example refers to the hardware implementation of the RSA algorithm and to the elliptic-curve public key scheme. They typically scan the bits of the private key in a serial fashion. If the required time for implementing the different algorithm functions is not the same than these timing differences leak information about the key. For example, it is crucial to guarantee that the finite-state machine (FSM) expend the same number of cycles for each operation regardless of the processed data. Similarly, a correlation between the IC power consumption or electromagnetic radiation can reveal sensitive data. Unlike the timing attacks which can be executed remotely, power attacks require to the attacker of being physically close to the device. These attacks are passive since they monitor the normal operation of the device without disturbing it. More precisely, simple power attacks (SPAs) rely on few power or EM measurements for extracting sensitive information. An example are the template attacks in which a huge number of measured data is required for creating the template but then it requires few measurements for being implemented. On the other hand, differential power attacks (DPAs) require multiple power or EM traces for being implemented. Indeed, the attacker creates a model of the power consumption profile of the circuit and assumes that the power consumption is related to the Hamming distance (HD) between current and previous data in registers or flip-flops. Typically, these attacks are used to reduce the computational complexity of the brute-force attacks [3]. Finally, fault attacks cause faulty errors into an IC by playing with the power supply or, for example, by inducing clock glitches [4]. Some of the most used countermeasures are:

- *Leakage reduction*, which decreases the dependency between side-channel traces and the key information. Unfortunately, this approach does not eliminate the criticalities of the attacks. Indeed, the side-channel information are strictly correlated to the system's input in the CMOS technology. Anyway, several leakage reduction techniques allow reducing the impact of these attacks. Some of them are smoothing the power consumption by using differential logic, current-mode logic or dual-rail with pre-charge logic.
- *Noise injection*, which helps to reduce the signal-to-noise ratio (SNR) of the side-channel information. This approach does not eliminate the problem but increase the required work of an attacker for disclosing sensitive data. Indeed, as mentioned above, DPAs require a huge number of measurements for performing the attack thus intrinsically reducing the impact of the noise.
- *Key update*, which refers to frequently update the secret key for preventing the accumulation of side-channel information by the adversary.
- *Side-channel-resistant PUFs*, which mean the use of auxiliary circuits able to reduce the impact of side-channel attacks on these primitives.

- *Secure scan chains*, which refer to the reduction of sensitive registers in the scan chains or at least to the protection of these registers through several techniques which involve, for example, the use of mirror key registers.

Finally, Counterfeiting consists of imitating or damaging an IC with the aim of stealing the IP or harming the reputation of the authentic provider. Several techniques have been developed for detecting this attack such as:

- *Hardware metering*, which consists of tracking the ICs through a set of tools and methodologies. This approach can be implemented in an active or passive way. The first one refers to the locking of some functionalities in the IC making them only accessible by the designer. On the other hand, passive metering involves to in an identification of particular IC functionalities and used for metering.
- *PUFs* as IC fingerprint.
- *Device aging*, which refers to monitoring the IC lifetime which is influenced by phenomena such as negative temperature bias instability (NBTI), hot carrier injection (HCI), and electron migration for avoiding someone sells a used IC as new.
- *IP watermarking*.

Depending on the algorithm topology the circuit optimization can be different [5]-[12]. In particular, the critical part when developing circuits for DES and AES implementations is related to the substitution boxes (i.e., Sboxes). In this case, special effort is expended to make them fast and compact. On the other hand, for applications which involve the public key algorithms such as RSA or elliptic curve-based cryptography, much effort is required for optimizing the hardware implementation in terms of small area, high throughput, low power, low energy, paying particular attention to not include sensitive registers (such as key registers) on the scan chains.

1.2 Motivations

Preserving information security in IoT systems is becoming a crucial issue. The development of the IoT network with billions of distributed electronic systems introduces several hardware vulnerabilities along with the benefits thus pushing the demand of preserving sensitive and secret data down to the chip level. Most of the security algorithms and protocols require of using a secret key as a root of trust. In particular, this key must be a truly random entropy source in a deterministic way. Typically, it is generated off-chip and stored in a non-volatile memory (NVM) but unfortunately this approach requires additional costs and suffers from hardware attacks. Indeed, it requires circuits always powered on for protecting the key implying a higher energy budget thus complicating the IC design. Hardware primitives such as PUFs represent emerging solutions which exploit truly random physical phenomena for generating a unique, repeatable, and secure key in a volatile fashion. However, ensuring an adequate PUF stability (i.e., repeatability) is still a challenge thus requiring stability enhancement techniques which result in lower area and energy efficiency.

This thesis aims to introduce a novel class of CMOS PUF for hardware security applications. The purpose is of exploiting the variability in CMOS manufacturing processes as static entropy source for generating a deterministic truly random number with high reliability to process, voltage and temperature (PVT) variations thus reducing the need of stability enhancement techniques which

degrade the area and energy efficiency. Moreover, they also introduce more hardware vulnerabilities such as helper data used for recovering the key that need to be stored in an NVM.

1.3 Thesis overview

This thesis is organized in six chapters. Following this introduction, chapter 2 provides the background and shows the state of the art on the physical unclonable functions. Chapter 3 discusses the proposed PUF solution in 180-nm CMOS technology. Chapter 4 reports the simulation results obtained by simulating the proposed PUF along with other relevant solutions in an 2D technology. Chapter 5 discusses a possible application scenario, using the proposed PUF solution as building block for implementing a smart tag. Finally, chapter 6 concludes this thesis. More in detail:

Chapter 2 provides a general overview on the field of PUFs, with a particular attention on silicon PUFs. More precisely, this chapter starts talking about the process variations in CMOS manufacturing process and how they can be exploited for generating a unique ID. Later, the most important PUF metrics are reported at which it follows a small overview on the main possible applications. Finally, a perspective on the most relevant works is provided at the end of this chapter.

Chapter 3 introduces the class of static monostable PUFs based on a subthreshold voltage divider between two nominally identical sub-circuits. More precisely, this chapter starts with a general discussion on the adoption of a voltage divider as PUF core circuit at which it follows a complete description of different circuital variants analyzed during my PhD, supported by both simulations, measurement, and analytical equations. The main contents of this chapter are taken from our journal and conference papers: "*Static CMOS Physically Unclonable Functions Based on 4T Voltage Divider With 0.6%−1.5% Bit Instability at 0.4−1.8 V Operation in 180 nm*", published in *IEEE Journal of Solid-State Circuits* (*JSSC*) 2022 [13], and "*Stability-Area Trade-off in Static CMOS PUF Based on 4T Subthreshold Voltage Divider*" presented at the *IEEE International Conference on Electronics Circuits and Systems* (*ICECS*) 2022 [14].

Chapter 4 explores the possibility of using emerging devices such as paper-based MoS_2-FET for implementing PUF circuits. In particular, this chapter first provides a briefly introduction on 2D electronics and describes the MoS_2 FET fabricated on paper substrate [15]. Later, a description of how experimental results, detailed in [15], were exploited to setup a LUT-based Verilog-A model. Finally, this chapter reports simulations results of the proposed PUF circuit, implemented with these emerging devices. The main contents of this chapter are taken from our journal papers: "*Assessment of 2D-FET Based Digital and Analog Circuits on Paper",* published in *Solid-State Electronics* (*SSE*) 2021 [16], and "*Assessment of Paper Based MoS_2 FET for Physically Unclonable Functions"* published on *Solid-State Electronics* (*SSE*) 2022 [17].

Chapter 5 exploits the proposed PUF solution for implementing a smart tag. This chapter starts stressing how important is nowadays pushing the information security down to the chip level. Then it illustrates the proposed passive tag architecture along with an analysis against possible hardware and software threats. The main contents of this

chapter are taken from our paper "*PUF-Based Authentication-Oriented Architecture for Identification Tags*" submitted to *IEEE Transactions on Dependable and Secure Computing.*

Chapter 6 concludes this thesis with a summary of the obtained results and an overview of the future direction of these PUFs.

Chapter 2
PUF theory and applications

2.1 Introduction

Nowadays, the on-chip availability of secret and deterministic keys is becoming ever more crucial for guarantying information security [18]. Indeed, as discussed in chapter 1, several hardware attack topologies can be faced by using chip ID. Conventionally, secret keys are generated off-chip and stored in a non-volatile manner [19]. In particular, the most common used storage mediums are the one-time programmable (OTP) memories, where a fuse o anti-fuse is used for locking the bits (i.e., in this case data are written during the chip manufacturing process and cannot be changed), and non-volatile memories (NVMs) like Flash, FRAM and NRAM. Unfortunately, the OTP approach requires additional cost and expose the key to security risks since in the most of cases the devices are fabricated by a third-party facility which is not always a trusted-party [20]. On the other hand, NVM approach suffers from software attacks such as read-out attacks (i.e., a malware can gain an unauthorized access to the memory) and hardware attacks such side-channel and reverse engineering attacks (e.g., a malicious user can extract information on the secret key by analyzing the power consumption profile, the timing required for reading each bit, data remanence, etc.) thus requiring additional always-powered circuits for protecting the secret key. This also leads to additional energy costs which not always meet IoT constraints since these devices operate with battery or harvested energy. Ideally, the two following concepts must be ensured for guarantying security to a key embedded in an integrated circuit (IC) [21]: (*i*) secret key should not be vulnerable to physical inspections like imaging, reverse engineering and side-channel attacks; (*ii*) the key should be physically available only when the chip is powered on for reducing the vulnerabilities related to the disabled protection techniques when the chip is powered off.

In the last years, physically unclonable functions (PUFs) have been extensively explored to overcome these challenges. From a more general point of view, PUFs exploit truly random but deterministic physical phenomena as static entropy source for generating a unique, repeatable, and secure key in a non-volatile manner [22].

2.1.1 Chapter organization

The chapter is organized as follow. Section 2.2 gives an overview of the process variations in CMOS manufacturing processes. Section 2.3 describes the most important PUF metrics. Section 2.4 illustrates some possible application scenario. Section 2.5 and 2.6 provide an overview of the most relevant weak and strong PUF implementations. Section 2.7 discusses the most used stabilization techniques. Finally, Section 2.8 concludes this chapter.

2.2 Process variations

Among the different topologies, silicon PUFs leverage on the physical disorder inherent in the CMOS manufacturing processes among ICs with identical masks for uniquely characterizing each chip. Physical disorder refers to the random imperfections in the structure of physical objects [23]. These variations typically represent a negative effect for a designer. Indeed, several techniques have been developed from designer and manufactures for reducing the impact and the entity of these variations so that both designing and manufacturing phase must be optimized for improving the yield. However, despite chips passe the yield tests as if they are the same at the macro-level, it is impossible to find two chips with perfectly identical behavior when observing minor differences and it is expected to get worse when scaling the technological node since it is becoming more and more difficult to fabricate perfectly sized devices [24], due to the limitations imposed by quantum mechanics. Physical sources of variability can be categorized as follow:

- ***Geometry of the device,*** which includes the film thickness variations and the lateral dimension variations. The former refers to the variations of the gate oxide thickness (i.e., T_{OX}). On the other hand, lateral dimensions variations such as channel length (i.e., L_{eff}) and channel width (i.e., W_{eff}) are mainly due to photolithography proximity effects or plasma etch dependencies. MOSFET are particularly sensitive to L_{eff} and T_{OX} variations, since they directly affect the output current characteristics, and less sensitive to the W_{eff} variations. Actually, T_{OX} is a well-controlled parameter, indeed the biggest variations tend to occur mainly from one wafer to another wafer, as opposed to the L_{eff} which is still a critical parameter.
- ***Material of the Device,*** which refers to the internal material parameters such as doping and additional material (e.g., related to the deposition and anneal pahses). Doping variations are due to dose, energy, angle, or other ion implant dependencies and mainly affect the matching between nMOS and pMOS devices even when the variations in the same wafer and in the same die are very small. Deposition and anneal processes directly impact on the deviation of additional parameters. These are mainly observed in silicide formation and in the grain structure of poly or metal lines. The variation of these material parameters contributes to the contact and line resistance variation.
- ***Geometry of the interconnect line,*** which includes geometrical parameter such as line width, line space, metal thickness and dielectric heigh. Line width (i.e., w) and line space (i.e., s) variations in the patterned lines are mainly due to the photolithography and etch dependencies and primarily impact the line resistance and the inter-layer capacitances. Metal thickness (i.e., t) variations do not represent a critical parameter in conventional metal interconnect lines (i.e., this parameter mainly varies from wafer to wafer), since the deposited metal films is a well-controlled process. On the other hand, in the damascene processes (e.g., copper polishing) the dishing and erosion procedures can strongly impact the final thickness of the patterned lines thus resulting in large variations of the metal thickness within the wafer. Finally, dielectric height (h) variations refer to the thickness variations of the oxide films. This is primarily due to the deposition and polishing phases which contribute to the wafer level variation. Furthermore, chemical

mechanical planarization (CMP) process also introduces variations at the die level thus resulting to different oxide films height within the die.

- ***Material of the interconnect line,*** which refers to the variation in terms of metal resistivity, dielectric constant and contact and via resistance. Metal resistivity (i.e., ρ) is a well well-controlled parameter and typically varies from wafer to wafer. Dielectric constant (i.e., ε) is also a well-controlled parameter and the small observed variations are mainly due to the deposition process. Finally, contact and via resistance variations are caused by the clean and etch processes and mainly vary from wafer to wafer.

These variability sources lead to always have a different behavior even when the same circuit is implemented in different chips. Furthermore, the impact is expected to be much higher in future technologies (which ensemble heterogeneous structures with an even smaller sizing) thus making more and more harder to predict the chip performance in terms of power consumption, throughput and so on. This leads a digital and analog designer to always consider the worst-case scenario which can complicate the design phase. On the other hand, a PUF designer exploits these tiny differences in terms of device and interconnection materials and geometries for generating a unique fingerprint of a chip.

2.3 PUF metrics

The suitability of a PUF of being used for hardware security applications can be assessed by a set of well-established metrics such as ***randomness***, ***uniqueness***, ***reliability***, ***identifiability***, ***stability***, ***physical unclonability***, ***unpredictability***, and ***physical attack immunity***. Moreover, when we focus on silicon PUFs other important metrics should be considered such as ***area efficiency***, ***throughput***, and ***power and energy per bit***.

The importance of these metrics may be different depending on the application. For example, if PUFs are used for generating cryptokeys then uniqueness and randomness need to be ensured so that different devices show distinct derived keys. On the other hand, if PUFs are used for applications like low-cost authentication, the unpredictability needs to be optimized for reducing the possibility of an external model constructed by an eavesdropping attacker for predicting the other CRPs. Guarantying good performance in all the above metrics is not always possible thus highlighting the need of optimizing such metrics in relation to the application for which the PUF is intended. Moreover, trends in hardware security requires ever more compact designs with high power and energy efficiency. Indeed, if the targeted application is the IoT network it is important to meet the budget constraints in terms of energy and area.

2.3.1 Randomness

This metric ensures that the probability of having '1' (i.e., $\Pr(1)$) and '0' (i.e., $\Pr(0)$) is the same in the PUF response (i.e., $\Pr(1) = \Pr(0) = 0.5$) so that an adversary who is observing the output of the PUF cannot deduce information about the PUF behavior (i.e., there is no more efficient attack than the brute force attack). The commonly used approaches for assessing the randomness of a PUF instance are *uniformity*, *entropy*, *spatial correlation*, and *statistical tests*.

The former estimates the percentage of '1' (i.e., $\Pr[1]$) and '0' (i.e., $\Pr[0]$) in a PUF response.

It can be evaluated as follow:

$$Uniformity = \frac{1}{N_{bit} \cdot R}\left(\sum_{i=1}^{R} HW_i\right) \quad (2.1)$$

Where R and N_{bit} represent the number of PUF responses and their bitlength respectively and HW is the Hamming Weight of the PUF response (i.e., the number of bits which differs from '0'). Ideally, a number for being truly random requires the same percentage of '1' and '0' (i.e., $\Pr[1] = \Pr[0] = 0.5$).

Entropy refers to the amount of information carried by each bit [21] and ranges between 0 (i.e., each bit carries no information, and it is perfectly predictable) and 1 (i.e., each bit carries a full bit information, and it is not predictable). This parameter is used in cryptographic applications for quantifying how unpredictable is the PUF response. The required effort for successfully performing a brute force attack is proportional to the number of key combinations in the key space size which is also function of the entropy (i.e., $key\ space\ size = 2^{keylength \cdot entropy}$). Indeed, higher entropy will result in a higher complexity from an adversary of breaking the key. This parameter is strictly related to the probability of having a bit '0' and '1' in the PUF response (i.e., an entropy value of 1 refers to have $\Pr[1] = \Pr[0] = 0.5$). In a practical case, a good entropy will correspond to a loss of effective keylength (i.e., $keylength \cdot entropy$) lower than 1 bit so that the probability having a bit '0' is quite close to that of having a bit '1'. Typically, this parameter is assessed by using Shannon entropy or min-entropy approach. The former is commonly used in cryptographic applications and, in the case of binary response, can be expressed as follow:

$$Shannon\ Entropy = -[\Pr[0] \cdot log_2(\Pr[0]) + \Pr[1] \cdot log_2(\Pr[1])] \quad (2.2)$$

Where Pr [0] and Pr [1] refer to the probability of having a bit '0' and '1' in the PUF response respectively. On the other hand, min-entropy represents a more pessimistic notion of Shannon entropy. Indeed, it is defined as the probability of successful guess of the most likely key value [21] and refers to the worst entropy scenario.

$$min\text{-}entropy = -log_2[\max(\Pr[0], \Pr[1])] \quad (2.3)$$

Fig.2.1 shows the Shannon entropy and min-entropy trends a function of $\Pr[0]$ along with a numerical example for better understanding this concept. If we consider 256-bit words, entropy should be at least 0.996 for ensuring a degradation in the effective keylength lower than 1 bit. This implies that if we consider the Shannon entropy, we need to achieve a Pr [0] probability in the range of 0.463-0.537. On the other hand, if we consider the more stringent min-entropy the Pr [0] probability must be very close to 0.5 (i.e., between 0.498 and 0.502).

To make the produced bits hard for being predicted it is also important reducing the spatial correlation between neighboring bits. It is important avoiding layout dependent variations. To this purpose, autocorrelation function (ACF) can be used for estimating the spatial correlation between neighboring bits. This function aims to find similarity between observed random samples as function of the spatial between them and ranges between 0 (i.e., no spatial correlation exists between bits which are spatially close) and 1 (i.e., neighboring bits are correlated among them).

Finally, the randomness of PUF responses can be assessed through well-established statistical tests such as NIST (i.e., National Institute of Standards and Technology) test [25]. However, these

tests require a certain number of samples so that it might not be always possible relying on this approach with a high reliability.

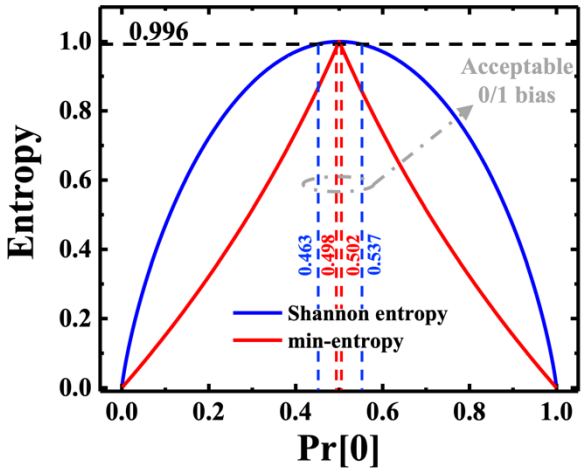


Fig. 2.1. Shannon entropy and min-entropy versus Pr[0].

2.3.2 Uniqueness

Another important PUF feature is the ability of generating a unique response like a digital fingerprint. PUF instances must show a distinguishable behavior when compared with the same PUF instances implemented in other chips. This metric is evaluated by using the inter-chip Hamming Distance (i.e., HD_{inter}) whose value should be as close as possible to the 50 % thus indicating that each PUF instance shows a unique behavior when compared to the same PUF solution implemented in different chips. Indeed, considering i and j (with $i \neq j$) as two different chips with N_{bit} responses R_i and R_j for a given challenge the uniqueness can be expressed as follow [23]:

$$HD_{inter} = \frac{2}{N_{chip} \cdot (N_{chip} - 1)} \sum_{i=0}^{N_{chip}-1} \sum_{j=i+1}^{N_{chip}} \frac{HD(R_i, R_j)}{N_{bit}} \quad (2.4)$$

Where N_{chip} refers to the number of chips under test and $HD(R_i, R_j)$ refers to the Hamming distance (i.e., the number of positions where they differ) between the two chosen responses. For better clarifying this concept we can consider the example in Fig. 2.2 which provides an example on how evaluating the uniqueness between two identical PUFs implemented in two different chips.

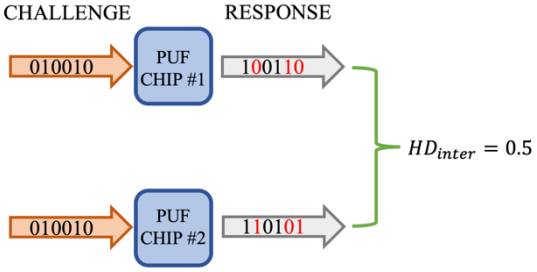


Fig. 2.2. An example of how evaluating the HD_{inter} between two PUF instances.

From this figure, for a given challenge, the two 6-bit responses show 3 different bits from each other thus resulting in a 0.5 Hamming Distance.

2.3.3 Reliability

Reliability measures how consistent is the PUF response (i.e., R) for a given challenge (i.e., C) regardless of the noise or different environmental conditions. Ideally, the PUF response should be the same even under noisy conditions and voltage and temperature variations. This parameter can be evaluated by performing the intra Hamming Distance (i.e., HD_{intra}) between the PUF responses, for a given challenge, achieved under noisy or different environmental conditions. Indeed, considering a $i-$chip with N_{resp} responses achieved under different environmental conditions the HD_{intra} can be calculated as follow [23]:

$$HD_{intra} = \frac{1}{N_{chip}} \sum_{i=1}^{N_{chip}} \frac{HD(R_i, R_i')}{N_{bit}} \quad (2.5)$$

Where R_i and R_i' refer to the N_{bit} responses achieved at nominal (i.e., golden key, GK, conditions) and different environmental conditions respectively. From this equation we can write:

$$Reliability = 1 - HD_{intra} \quad (2.6)$$

This parameter varies from 0 to 1, where 0 indicates that the PUF instance is not reliable to noise or different environmental conditions while 1 indicates that the PUF response is consistent regardless to the environmental conditions. Fig. 2.3 illustrates an example of how evaluating the PUF reliability under temperature variations.

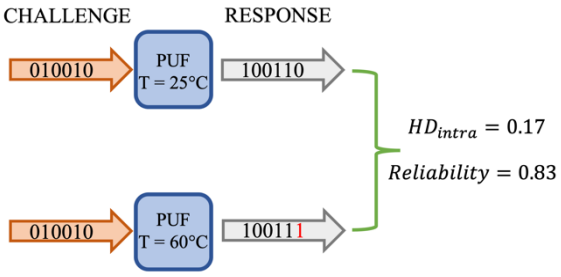


Fig. 2.3. An example of how evaluating the PUF reliability through two responses obtained at different temperatures.

This figure provides two responses of the same PUF obtained with the same challenge but under different environmental conditions. From this figure the two 6-bit responses differ of 1 bit from each other thus resulting in a reliability of 0.83.

2.3.4 Identifiability

Identifiability measures the PUF ability of showing a distinguishable behavior even under noisy or different environmental conditions [21]. It is related to both uniqueness and reliability and can be expressed as follow.

$$Identifiability = \frac{HD_{inter}}{HD_{intra}} \quad (2.7)$$

Ideally, an identifiable PUF instance should deliver a unique (i.e., $HD_{inter} = 0.5$) and deterministic (i.e., $HD_{intra} = 0$) response at all the considered conditions.

2.3.5 Stability

Ideally, PUFs represent circuit solutions which perfectly exploit within-die variations for generating secret keys or IDs while rejecting, at the same time, the effect of all other variations [21] such as:

- *Die-to-die variations*, which indicate that the PUF repeatability should not be affected by die-to-die variations (i.e., systematic process variations).
- *Environmental variations*, from which the PUF instance should deliver a consistent response regardless to the inevitable voltage and temperature variations.
- *Aging,* which implies that the PUF response should be consistent during the overall lifetime of the device.

Stability represents a very crucial issue since in many cryptographic protocols one-bit change results in a completely different cipher text. Nowadays, keeping low the instability is one of the major challenges for PUF designers. Anyway, it can be estimated through few important metrics such as unstable bits, bit error rate (BER), key error rate (KER), and mean time before failure (MTBF). The former refers to the cumulative count of flipping bits under different evaluations over the entire population of cells under noisy or different environmental conditions. In particular, the unstable bits include noisy bits (i.e., bits which flip at least once under different evaluations due to on-chip noise) and flipped bits (i.e., bits which permanently flip when changing the environmental conditions compared to the Golden key). On the other hand, BER counts the average of the simultaneous instability exhibited by the PUF output word [21]. It is also strictly connected to the KER (i.e., the probability of having at least one flipped bit in the PUF response) as follow:

$$KER = 1 - \sum_{i=0}^{K} \left(\frac{N_{key}}{i}\right)(1 - BER)^{N_{key}-i} BER^{i} \quad (2.8)$$

Where K and N_{key} represent the maximum number of bits potentially corrected by an ECC hgand the key length respectively. This parameter must be kept low (i.e., typically 10^{-6}) so that the MTBF is equal to, or at least comparable to, the life of the device. In particular, the MTBF refers to the ratio between the average inter-access time (i.e., $t_{inter-access}$), between two consecutive PUF accesses [21], and the KER as follow.

$$MTBF = \frac{t_{inter-access}}{KER} \quad (2.9)$$

For example, if we consider a duty-cycled sensor node which sends measurements every time it is woken up, the $t_{inter-access}$ refers to the following time between two successive wakes up events. Indeed, the targeted KER should be set according to the following time between two consecutive accesses which is strictly related to the intended application of the device. At the same time, the BER needs to be properly kept low for reaching the targeted KER.

2.3.6 Physical unclonability

Physical unclonability refers to the ability for a PUF instance of being always distinguishable from its clones. Considering an authentic PUF instance I_A and its clone I_C with their respective CRP space. For a given set of challenges C the two instances I_A and I_C will generate the two set of responses R_A and R_C respectively. The authentic instance for being physical unclonable should exhibit an average HD between the elements of R_A and their corresponding (i.e., delivered with the same challenges) counterparts in R_C much larger than the average HD_{intra}, evaluated at different environmental or noise conditions. This can be mathematically expressed as follow.

$$\frac{1}{C}\sum_{c\in C} HD(R_A, R_C) \gg HD_{intra} \quad (2.10)$$

This means that the PUF behavior should be distinguishable from other clones even under different environmental or noise conditions.

2.3.7 Unpredictability

Unpredictability (i.e., mathematical unclonability) refers to the ability of a PUF instance of showing a distinguishable behavior from the any PUF model built by an adversary. Considering the authentic PUF instance I_A and another implemented with a mathematical model I_M (i.e., supposing that the adversary has access to a significant number C_M of CRPs with which he can build a model) with their respective CRP space composed by a set of challenges C and the respective set of responses R_A and R_M respectively. The authentic instance can be defined unpredictable if the average HD between any element of the two set of responses R_A and R_M is much larger than the average HD_{intra} between the elements in R_A evaluated at different environmental and noise conditions. This concept can be mathematically written as follow.

$$\frac{1}{C_M}\sum_{c\in C_M} HD(R_A, R_M) \gg HD_{intra} \quad (2.11)$$

This means that the average error (i.e., the HD) produced by the model must be significantly higher than the error due to different environmental and noise conditions. Unpredictability can be estimated by using different techniques such as:

- *conditional entropy,* which estimates the minimum bit number that cannot be predicted by an adversary which knows a certain number of CRPs
- *machine learning algorithms*, where an adversary uses a set of CRPs for training a software model of the PUF which is then validated using the remainder of the CRPs).
- *HD test*, which estimates the output transaction probability of a PUF.

2.3.8 Physical attack immunity

The big impact of implementing a PUF solution as alternative to NVM based approach relies on the native resilience to the hardware attacks since the secret key is generated on the fly instead of being stored in a non-volatile manner. Nowadays, however, increasingly effective attacks have

been developed for leaking information on the secret key. For this reason, testing the proposed solutions under different attack topologies is becoming a crucial issue.

2.4 PUF applications

Ideally, PUFs can be seen as digital blocks that respond to inputs (challenges) with repeatable outputs (responses) thus generating a challenge response pair (CRP) in an unpredictable way. The latter property refers to the fact that the input-output mapping is unknown to an external observer. Moreover, the responses are defined by chip-specific random variations and are generated on the fly thus requiring the chip of being powered on for the deployment of the keys. PUFs can be grouped in weak and strong based on the number of generable CRPs [21]. Weak PUFs are categorized by a number of CRPs which increases linearly with the physical implemented bitcells. The poor capability of CRP space makes these PUFs suitable for being used as cryptokeys instead of the disclosure in insecure channels. On the other hand, strong PUFs exhibits a number of CRPs which increases exponentially with silicon implemented bitcells. The large capability of the CRP plan allows in-plain transmission because replay attacks are counteracted by the very low probability of reusing the same CRP. The difference in terms of generable CRPs makes the two PUF classes suitable for different applications.

2.4.1 Cryptographic key generation

Nowadays, ensuring security to electronic devices which deal with sensitive and private information is a required crucial task. These systems should be able to protect data, verify information integrity and execute other security functions. Such requirements are typically achieved by using encryption algorithms and hash functions. These blocks rely on a secret key that should be known by only trusted users. Typically, weak PUFs are used for generating cryptographic keys due to a lower capability of generating CRPs. For this application topology these PUFs need to satisfy the following requirements:

- ***High reliability,*** which implies that the PUF instances must deliver the same response, for a given challenge, even under noisy or different environmental (i.e., voltage and temperature variations) conditions. This is because even a single bit change would completely disrupt the couple plain/cipher text [26] thus making the decryption very difficult. For this reason, the PUF native response needs to be post processed with additional circuits.
- ***Uniqueness***, which refers to the fact that each key should be unique with respect to the other keys generated in different chip (i.e., electronic systems) so that if one key is compromised the others remain secure.
- ***Randomness***, which increases the difficulty of implementing brute-force attacks for guessing the key. When the PUF response is not uniformly distributed additional circuits are required for compressing enough entropy in a PUF-generated key.

The process which generates a cryptographic key from PUF instances can be divided in two macro steps [23]: setup stage and key generation summarized in Fig. 2.1 (a) and (b) respectively.

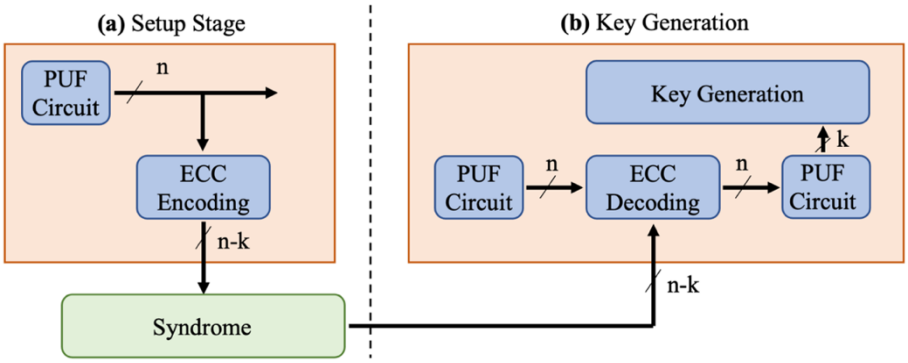


Fig. 2.4. Cryptokey generation procedure: (a) setup stage and (b) key generation.

The setup stage is implemented only once by the developer and includes: (*i*) pre-processing phase for estimating the maximum BER under both noisy and different environmental conditions (i.e., at the design stage different test chips are used for evaluating the reliability of the PUF instances); (*ii*) helper data generation phase for generating public information (i.e., syndrome) used for correcting any occurred bit flip in the PUF response (i.e., this information can be stored anywhere even with bit vector for selecting the pairs); (*iii*) device enrolment for which the keys generated by each device are stored securely by an authentication authority for ensuring secure communication. Obviously, syndrome information represents a vulnerability point for an attacker who can use it for guessing the key. However, using a b-bit of syndrome an attacker can guess at most b bits of the PUF response. Therefore, to obtain k secret bits we can generate n = k + b bits from the PUF circuit so that even with the syndrome information, an adversary needs to guess at least k bits. An example is the Bose-Chaudhuri-Hocquenghem **(**BCH) code which represent a special class of cyclic codes with the ability to correct more than one error. Indeed it can be written as BCH(n, k, d) and represents an error-correction code able to correct up to (d-1)/2 errors out of n bits with an (n-k) bits of syndrome (i.e., b = n – k). The second step is the key generation (i.e., whenever the key is required) and includes the following procedures: (*i*) stable response construction which refers to feed the PUF response into an error correction block for re-generating a reliable response (i.e., supported by the helper data); (*ii*) privacy amplification which consists of applying the re-generated response to an entropy compression block (e.g., hash function) for enhancing the randomness; (*iii*) key derivations of single or multiple keys for different security tasks (e.g., encryption, identification, etc.) by using the output of the entropy compression block; (*iv*) After the previous procedures the PUF instance is powered off so that the key is no longer accessible. In this way, the PUF responses can generate keys for any cryptographic operations. Indeed, for cryptographic operations the ECC output can be hashed down to a desired length and used as a cryptographic key (e.g., symmetric key primitives such as AES can used the hashed PUF output). For cryptographic operations where the key must satisfy some property the hashed PUF output is used as a seed for the key generation algorithm (e.g., algorithms like RSA require key having specific mathematical properties).

2.4.2 Low-cost authentication

The identity of a physical object needs to be identified before a service can be offered. We can see the authentication authority as the verifier and entity of the physical object as the prover. For

example, the government is the verifier of e-passport while the bank is verifier of the credit cards. Typically, in these applications when an entity (i.e., prover) want to authenticate itself to a verifier, it should provide evidence of its entity (i.e., generated by the entity itself) and a proof that the entity is actively involved at the time of authentication with the aim of convincing the verifier that it has exclusive access to secret and sensitive information. Generally, this is achieved in two steps: (*i*) identity provisioning which refers at the phase in which each device receives an unique identity and (*ii*) verification phase where that identity is required by the verifier for validating the identity of each entity. A conventional approach is the ISO/IEC 9798-2 standard, which uses the symmetric challenge-response technique. In this standard, during the provisioning phase:

1) Verifier gives to each entity a unique secret key (k) and a unique identifier (ID).
2) Verifier stores this information in a database.

During the verification phase:

1) Prover which wants to authenticate itself sends the ID to the verifier.
2) Verifier control the k associated to that ID.
3) Verifier sends a random number (once) to the prover.
4) Entity encrypts the nonce with the secret key and sends response (ne) back to the verifier.
5) Verifier decrypts the received response (ne) using the k associated to that entity.
6) If the decrypted data match the nonce sent by the verifier, the prover can be authenticated.

This approach presents two major disadvantages: (i) the provisioning phase requires to assign to each entity a secret key and this should be done during fabrication phase thus suffering from the same problem reported before (e.g., the possibility of having an untrusted third-party facility). (ii) each entity should implement an encryption algorithm or keyed hash function, and this could be unaffordable in resource-constrained devices such IoT devices or RFID.

Strong PUFs can overcome to these disadvantages by generating unique keys for each device on the fly. Moreover, they also do not require the implementation of the encryption algorithms since their large capability of generating CRPs allows of using each of them only once thus cancelling out the effect of the replay attacks.

The core principle is to exploit the CRPs provided by a PUF instance for generating an inherent identifier for each physical entity. The PUFs used for this application should possess some quality such as:

- ***Mathematical unclonability*** which ensure that an adversary that is running an eavesdropping attack cannot guess the CRPs by building a software clone of the device.
- ***High reliability*** which means that the generated CRPs must be the same under different evaluations regardless of noisy or different environmental conditions (i.e., bit flip can lead to a denial of service).
- ***Uniqueness*** which ensures that the CRP behavior is unique for each device in the network so that each of them can be easily identified.

These properties can be more relaxed when using different authentication protocols. For example, when the number of used CRPs is limited, it is not possible for an adversary to construct a PUF model by using machine learning algorithms. On the other hand, some protocol can tolerate instability of a few numbers of bits (e.g., by associating the correctness of a response to the threshold proximity to the golden value rather than to bit accurate matching) thus relaxing the reliability requirements.

The most common protocol is the unilateral authentication scheme where a central authority acts as the verifier while the distributed devices which embed a PUF instance act as prover. During the years, different versions have been proposed but the basic operative principle is summarized in Fig. 2.2 and consists of enrolment and verification steps [26].

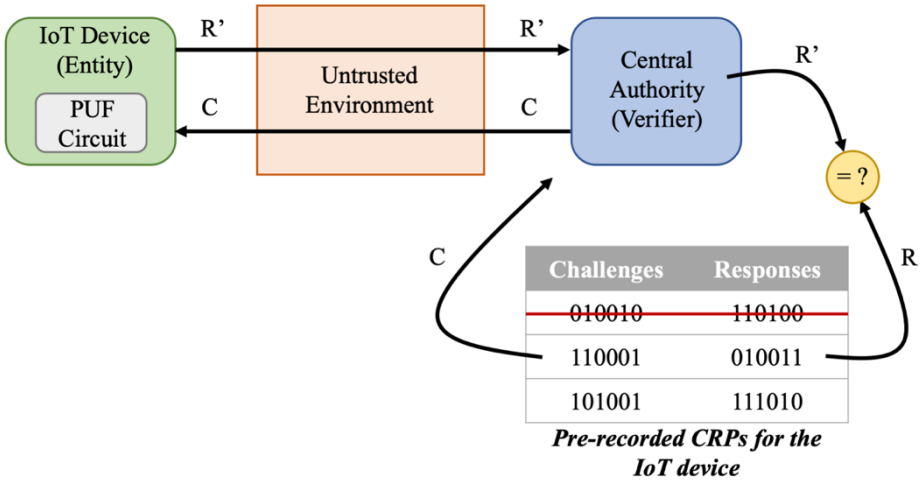


Fig. 2.5. Operative principle of PUF based authentication process.

During the enrolment phase:

1) Verifier (or a trusted third-party) embeds a PUF instance in each entity device and gives to them unique IDs.
2) Verifier provides a huge number of challenges (C) to the devices and records the corresponding responses (R).
3) Verifier builds up a secure database in which he can store all the IDs with the corresponding recorded CRPs.

In the verification phase:

1) Entity who wants to authenticate itself sends the ID to the verifier.
2) Verifier controls the CRPs associated to that ID.
3) Verifier take one of the stored challenges (C) and sends it to the entity.
4) Entity applies the challenge (C) to its PUF instance and sends the response (R') back to the verifier.
5) Verifier compares the received response (R') with that recorded in the database (R) associated to that challenge (C). If these two responses match the entity can be successfully authenticated.
6) Verifier deletes the CRP used in the previous authentication for preventing replay attacks.

However, this basic approach presents two major drawbacks: (*i*) each device needs to be enrolled and this is not a scalable process if we think to a modern IoT network composed by billions of distributed electronic devices; (*ii*) this approach is sensitive to machine learning modeling attacks since an eavesdropping user can collect enough CRPs for building an adequate model of the PUF (this can be solved by using obfuscation circuitry for, per example, permuting the input challenge).

Slender PUF was proposed to overcome to these issues [23]. During the enrolment phase:

1) Verifier (or a trusted third-party) embeds a PUF instance in each entity device and gives to them unique IDs.

2) Verifier provides a huge number of challenges (C) to the devices and records the respective responses. Later it constructs a mathematical model for each PUF using machine learning algorithms.
3) Verifier stores all the IDs with the corresponding software model.

During the verification phase:

1) Entity who wants to authenticate itself sends the ID and a random binary vector (nonce e) to the verifier.
2) Verifiers checks the ID and send to the entity another random binary vector (nonce v) then they both concatenate the two nonce for generating (e, v).
3) Entity use a pseudorandom function G (i.e., previously concorded) for constructing a challenge c based on the seed. This challenge is later applied to its PUF for generating a response r of m bits.
4) Verifier used the same pseudorandom function for constructing the same challenge based on the seed and provides it to its PUF model associated to that ID for generating the response r' of m bits.
5) Entity sends a sub-string s of the response back to the verifier along with the relative indexes.
6) Verifier performs the HD between the received sub-string s and the generated one s' with the model (with the same indexes provided by the entity) and if the results is smaller than a threshold t, the entity is successfully authenticated.

This protocol increases the complexity by an eavesdropping adversary of modeling the PUF instance since the challenge is not transmitted in the channel. In other advanced protocol [23] a partial challenge is sent to the receiver for verifying that there is an embedded PUF instance in the device. This partial challenge is later padded with a random pattern generated by a pseudorandom function for achieving a full-length challenge before applying it to the PUF instance. In this case the verifier uses a challenge recovery mechanism for generating an emulate response to compare with the received one. Another approach consists of dividing the challenge in two sub-challenges: valid and invalid. The former is called secret-challenge and the number of these challenges is not sufficient for an eavesdropping adversary for building a mathematical model. Obviously, these advanced algorithms increase the require time and resources for implementing the protocol.

2.4.3 Hardware-assisted cryptographic protocols

Many applications such as data mining, electronic voting, and anonymous transactions require a secure multiparty computation where several parties carry out joint communication based on their private inputs. In this case the security requirements are: (*i*) no single party can know something about the private inputs of the other parties through the protocol; (*ii*) each individual inputs should be independent from the others; (*iii*) Only the authorized parties can access to the output of the protocol.

Implementing a protocol that meets these three requirements is not an easy task. Hardware-assisted cryptographic protocols implement tamper-proof hardware tokens for improving the security during a multiparty computation. Indeed, in these protocols the trust between the parties is established through the exchange of hardware tokens. Such few examples include government-

issued signature cards for generating private/public key pairs for digital signatures, smart card-based scheme in data mining or secure memory for limiting the number of accesses. The unpredictable CRP behavior showed by the PUF instances can be exploited for implementing hardware-assisted cryptographic protocols such as: key exchange (KE), oblivious transfer (OT) and bit commitment (BC).

Key exchange protocols are used when a secret key needs to be shared between two or more parties for initializing a secure communication. To better understand how these protocols work let's suppose that Bob and Alice want to communicate in a secure way [23]:

1) Bob applies two challenges (c1 and c2) to its PUF and achieves two responses (r1 and r2).
2) Bob sends its PUF physically to Alice.
3) Alice sends an acknowledge back to Bob for notifying him that PUF is arrived.
4) Bob sends pair (c1, r1) and c2 to Alice.
5) Alice applies c1 to the PUF and checks if the obtained r1' matches r1. If it is not the communication is terminated.
6) Alice applies c2 to the PUF for obtaining r2.
7) Bob and Alice use r2 for deriving a shared secret.

An oblivious transfer protocol enables a sender to send one or multiple data to a receiver while being oblivious to what items have been sent. One example is called 1-of-2 oblivious transfer, where Bob (i.e., the first party) retrieves one of two sent items without knowing anything on the other item. At the same way Alice (i.e., the second party) sent to Bob two items without having knowledges on what of the two items was retrieved by Bob. This protocol can be extended to k-of-n oblivious transfer and involves in some interesting application including zero-knowledge proofs and bit-commitment scheme. Strong PUFs can be used in this protocol since they exhibit a large number of CRPs. For explaining how strong PUFs are involved in this protocol we consider a 1-of-2 oblivious transfer between Bob and Alice [23].

At the beginning Alice holds two secrets $b_0, b_1 \in \{0,1\}^\gamma$ and Bob makes a choice $s \in \{0,1\}$. At the end of this protocol Bob will learn one of the two secrets held by Alice and she will not know anything about the Bob's choice. This protocol consists of the setup and execution phases. During the setup phase:

1) Bob (i.e., the receiver) applies a set of challenges $(c_0, c_1, \dots, c_k)$ to the PUF and collects and stores the responses $(r_0, r_1, \dots, r_k)$in a secure database. With $r_0, r_1, \dots, r_k \in \{0,1\}^\gamma$.
2) Bob gives its PUF to Alice (i.e., the sender).

During the execution phase:

1) Alice generates two random numbers (x_0, x_1) and sends them to Bob.
2) Bob takes a CRP (c, r) from the database and computes $v = (c \oplus x_s)$ sending it to Alice. For a sake of simplicity let's consider the Bob's choice was $s = 0$.
3) Alice computes the challenges $c_0 = v \oplus x_0$ and $c_1 = v \oplus x_1$ and applies them to the PUF instance then recording the two responses r_0 and r_1. In this case $c_0 = c \oplus x_0 \oplus x_0 = c$ and $c_1 = c \oplus x_1 \oplus x_0$.
4) Alice computes $r_0 \oplus b_0$ and $r_1 \oplus b_1$ and sends them to Bob.
5) Bob can finally deduce his chosen secret b_s so that $b_{s=0} = r_0 \oplus b_0 \oplus r = b_0$.

Note that $r_0 = r$ since both responses are linked to the same challenge c. Two important considerations need to be carried out: (*i*) the first assumption is related to the fact that both Bob

and Alice obtained from the PUF the same response for the same challenge thus implying a PUF reliability of the 100 %; (*ii*) Bob cannot know anything about b_1 since, with high probability he has not measured the pair (c_1, r_1). Indeed, the probability of having information about the other secret is function of the probability of recording the pair (c_1, r_1) and the probability of guessing c_1 if it exists in the database.

Bit-commitment scheme refers to a cryptographic protocol which allows one party (i.e., the commitment) to commit a chosen value while keeping it secret to the other party (i.e., the receiver). This scheme finds application in fields like verifiable secret sharing, secure billing protocols (e.g., where, for example, a consumer can prove to a provider his commitment to energy costs without revealing the actual value of the meter reading thus protecting consumer's detail). PUFs can be used for implementing this protocol by exploiting PUF-based OT protocols. In this case we need to invert the rules so that Bob (i.e., the receiver in the OT protocol) acts as BC-committer and Alice (i.e., the sender in the OT protocol) acts as BC-receiver. The protocol consists of commitment and reveal stage. During the commitment phase:

1) Bob (i.e., the BC-sender) acts as OT-receiver and uses a secret choice s, where $s \in \{0,1\}$, as an input of the PUF-OT protocol.
2) Alice (i.e., the BC-receiver) acts as OT-sender and uses her secrets b_0 and b_1, with $b_0, b_1 \in \{0,1\}^{\gamma}$,as input of the PUF-OT protocol.
3) The protocol works as an OT protocol where Bob is learning one of Alice'secrets.

During the reveal phase:

1) Bob sends the binary string composed by the pair s and b_s to Alice.

Substantially, if Bob can compute b_s his first choice must be s thus proving that his previous commitment to the secret choice was s.

2.4.4 Remote secure sensors

IoT network is composed by distributed electronic systems which embed wireless sensors for being immersed in an environment and performing sensing, transmission, and localized actuation. This opens to several applications such as environmental and structural monitoring, medical application and so on. Some of these applications require security protocols. Indeed, sensitive, and private data are transmitted during remote health monitoring. Existing solutions rely on using cryptographic block for encrypting and authenticating data and entities respectively. However, this solution suffers from two mainly issue: (*i*) data transferred from the sensing element to the cryptographic module are exposed to physical attacks; (*ii*) conventional approaches use classic cryptographic primitives such as symmetric ciphers or hash functions which could be prohibitive in terms of energy and costs for an IoT device.

Strong PUFs can effectively overcome to previous issues and being used for remote secure sensors since they do not require separate cryptographic module and represent a low-cost solution. As a main difference with respect to the other applications here we can exploit the PUF instability for remote sensing a physical quantity (PQ). In particular, PUFs are sensitive to the input challenge as well as the environmental variations. During the enrolment phase (i.e., before that PUFs are deployed in the different devices) PUF instances need to be tested at golden and different environmental conditions. In this way, for a given challenge the output response is

function of the environmental variations. These variations in the PUF response can be exploited for sensing a physical quantity.

The way in which this procedure can be effectively implemented is summarized in Fig. 2.3 where a transducer is used for translating the physical quantity into an electrical signal (i.e., in the figure a voltage signal). The latter can be used for modulating the PUF response so that at the receiver side the variations in the PUF response, for a given challenge, can be exploited for measuring the variations of the physical quantity. This solution does not require additional encryption stage for protecting data from eavesdropping attacks because only the parties who have access to the PUF can decrypt the response for achieving data information.

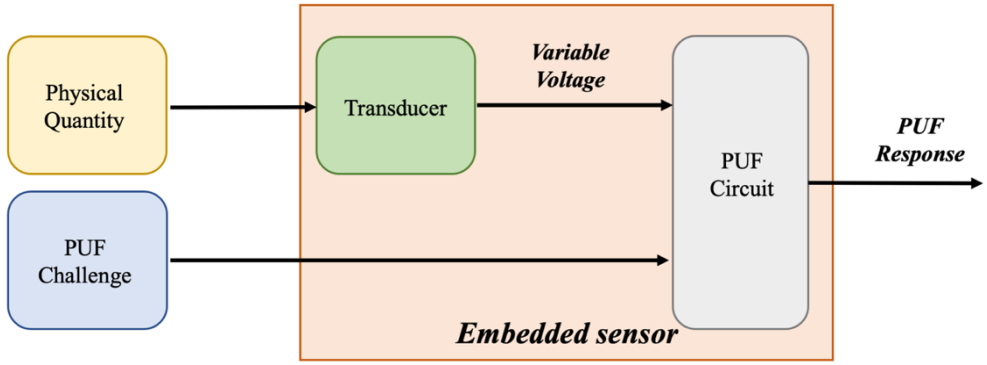


Fig. 2.6. PUF-based structure for remote secure sensing.

It is worth noting that for these applications it is better to use technologies which allow increasing the voltage range in which PUF can operate without being damaged. The sensor resolution is equal to the minimum voltage variation (i.e., ΔV) at which corresponds at least one-bit change in the PUF response. The implementation procedure is characterized by enrolment and sensing phases. During the enrolment phase:

1) A set of challenges is applied to each PUF instance for measuring the different responses.
2) The above measurements are repeated for k different voltages.
3) Sever must map the relation between the physical quantity to be measured and the voltage applied to the PUF instance.
4) For each PUF instance the server creates a database where CRPs at different voltages are stored along with the respective values of the physical quantity.

During the sensing phase:

1) The server applies a challenge (c) to the PUF instance.
2) The PUF generates a response (r) which is function of both applied challenge and voltage and send it back to the server.
3) The server checks in the database the value of the physical quantity associated to the received response.
4) The server cancels out the used CRP for protecting the device from replay attacks.

This protocol assumes that only the server has access to PUF characterization so that no encryption is used during the data transmission. Moreover, in the protocol described above we are neglecting that PUF could be sensitive to other environmental parameters such noise or temperature. Indeed, for being used for these applications the PUFs must be sensitive to the voltage variations but at the same time they should be very resilient to other environmental

variations like temperature variations or noise conditions. To assess how much suitable is the PUF for remote sensing applications we can use the following metrics:

- ***InterPQ distance*** ($\boldsymbol{HD_{InterPQ}}$) which refers to the HD between two responses to the same challenge at two different voltages (i.e., which differ by ΔV) but under the same noise conditions.
- ***IntraPQ distance*** ($\boldsymbol{HD_{IntraPQ}}$) which refers to the HD between two responses to the same challenge at different noise conditions but under the same voltage.

A strong PUF is suitable for operating under k different voltage levels if exist at least one challenge for which

$$min\left(\boldsymbol{HD_{InterPQ}}\right) > max\left(\boldsymbol{HD_{IntraPQ}}\right) \qquad (2.12)$$

With $min\left(\boldsymbol{HD_{InterPQ}}\right) > 1$. This condition means that the minimum HD between two responses obtained with the same challenge and under different voltages must be both higher than one and higher than the maximum number of occurred flips due to noise conditions. Another important feature for this application is the uniqueness (i.e., ideally equal to 50 %) so that if one device is compromised the others remain secure.

2.4.5 Anti-counterfeiting

The IC overproduction by malicious facilities is a real problem which causes significant financial losses to design houses. Weak PUFs can be effectively employed for implementing anti-counterfeiting mechanism. Indeed, PUFs can be embedded in a chip along with a proper locking mechanism during the design phase. In this way, the foundry applies the challenge chosen by the designer to each PUF instance and provides the obtained responses to the design house so that the designer authenticates each device and computes the passkey using the response to known challenge and sends it back to the foundry for chip testing purposes (i.e., without the passkeys the chips are locked). This process is sometimes referred to active hardware metering and allows the designer having more control over their designed chips. Indeed, only the authenticated chips will be used. This step will be better clarified with the following example summarized in Fig. 2.4.

1) The design house acts at register transfer level (RTL) level by adding a locking mechanism. This mechanism includes, for example, additional non-functional states in the finite state machine (FSM) so that the system is basically locked in one of these states and enters in a functional state only after a correct input sequence (i.e., the passkey). Moreover, adding additional states to the FSM design also helps for obfuscating the original functionality of the design.
2) The design house embeds the PUF circuit in each chip for initializing the internal flip-flops of the design. This increases the area overhead but helps the design house having more control on the post-fabrication phase thus limiting the chip overproduction.
3) The design house creates the GDSI file of the layout and sends it to the third-party facility for producing the chip along with a specific challenge for the PUF.
4) In the post-fabrication stage, the facility applies the received challenge to the PUF and sends the response back to the design house.

5) The design house computes a key (i.e., for unlocking the chip) from the received response and sends it back to the manufacturer.
6) The manufacturer powers-up the device and applies the received challenge to the PUF, set M = 1 and enables one clock cycle thus setting the design in a non-functional initial state (i.e., generated by the PUF).
7) The manufacturer now switches M to 0 and applies the passkey (i.e., provided by the design house) to the primary inputs and enables the clock thus driving the design in a functional state.
8) The chip is now ready for being tested.

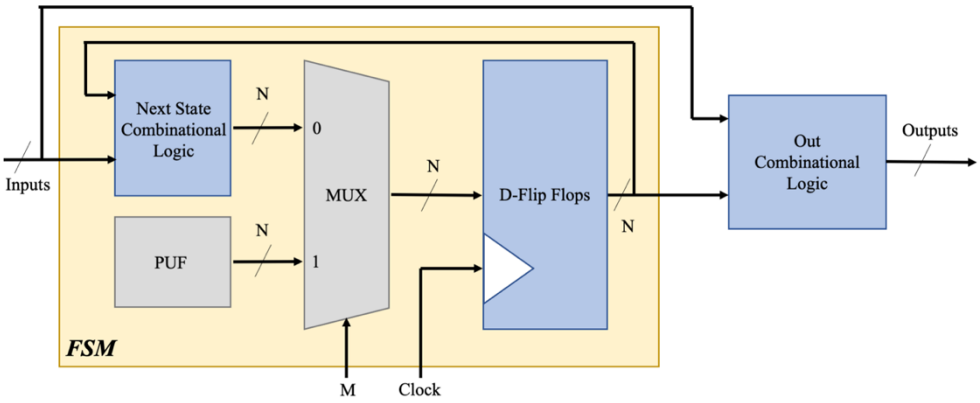


Fig. 2.7. FSM structure with embedded PUF for enabling locking mechanism.

The procedure described above needs to be implemented only once at the testing phase for protecting the design house against the overproduction. After this, the designer needs to let the system in a unlocked state. For doing this one possible solution consists of storing the passkey in a NVM so that the chip never goes back in a locked state. There is no need to encrypt the passkey such each device possesses its own key which is useless for other devices. It is also worth to point out that the approach described before assumes that the PUF circuit can generate a repeatable response regardless of the environmental or noisy conditions. This presupposes that the PUF stability must be improved even with stability enhancement techniques.

2.4.6 Tamper-proof design

PUFs open to several security applications which allow improving reliability to external hardware and software attacks. In the above discussion we only considered application related to silicon PUFs. Nowadays, other PUF topologies are designed and used for other different security applications. For example, another major problem in IoT network relies on the fact that billions distributed devices could be left unprotected in some specific environment thus being potential vulnerable to physical manumission. Here, Coating PUFs can be potentially used for implementing tamper-proof design where, laying out a network of metal wires as a comb shape produces a capacitor unique behavior for each device because, due to the manufacturing variations, metal lines show differences in terms of size or dielectric strength. This capacitor is then used as identifier for authenticating each device so that when an attacker tries to perform a

physical attack, he may modify the coating layer thus changing the value of the capacitor. This represents a low-cost solution for protecting IoT edge devices.

2.5 Weak PUF implementation

The huge impact of PUF concept for being implemented for hardware security applications pushed for designing ever more performing solutions. Indeed, in the last years several PUF topologies have been proposed and this number is expected to grow. Since the large number of existing solutions, it is not possible to well describe each of them. For this reason, this thesis mainly focusses on silicon based PUFs. Weak PUF refers to instances able to generate a number of CRPs that increases linearly with the number of physical implemented cells. Moreover, each generated bit is independent of each other (i.e., there is no statistical correlation between neighboring bits). Typically, PUF circuit is composed by the transformation block which transforms the static entropy source (such as the process variations inherently to the CMOS manufacturing process) into a measurable quantity (i.e., current, voltage, and delay) along with a conversion block which transforms this quantity in a binary response. Depending how this entropy source is translated into a measurable quantity we can categorize PUF solutions in SRAM and SRAM-based PUFs [27]-[35], delay-based PUFs [36]-[40], metastable-based PUFs [41]-[45], monostable-based PUFs [46]-[58], hybrid PUFs [59]-[61], active PUFs [62]-[66], Other PUFs [67]-[88]. Some of the most relevant solutions will be described below.

2.5.1 SRAM and SRAM-based PUFs

The SRAM PUF is one of the earliest proposed solutions for generating chip ID. In a more general case, this solution exploits the metastable behavior of the SRAM. Indeed, when powered up, this cell enters in a power struggle state then collapsing in one of the two possible stable states according to the transistor mismatch. In particular, the transistors belonging to the two cross-coupled inverters have identical nominal strength, but the random process variations ensure that one inverter has a stronger driving current than the other inverter thus pushing the cell in one of the two stable states. Two of the major advantages of using SRAM cells for PUF applications are: (i) the key can be generated by reusing the SRAM structures in the system thus reducing the area overhead for implementing this task; (ii) the differential read increase the reliability to power attacks since there is no difference when reading both the logical ‘0’and ‘1’. However, the conventional SRAM cells are very sensitive to noisy conditions and environmental variations. Indeed, during the power struggle state (i.e., the metastable state) if the mismatch is small, the on-chip noise can determinate the state in which the cell collapses thus resulting in a high percentage of unstable bits and native BER. For this reason, solutions which exploit the already existing SRAM cells employ additional circuits/algorithms for improving the native instability or generate the random bit by exploiting a different static entropy source. One of the most relevant SRAM-based PUF is reported in [29] and [30]. Here, the authors propose a new compact SRAM architecture with enhancement-enhancement structure (EE SRAM). This circuit aims to introduce a monostable state into the power-up behavior, which inherently improves the native stability

while keeping the other benefits of using a SRAM architecture. Fig. 2.8 shows the concept design along with the three possible working modes.

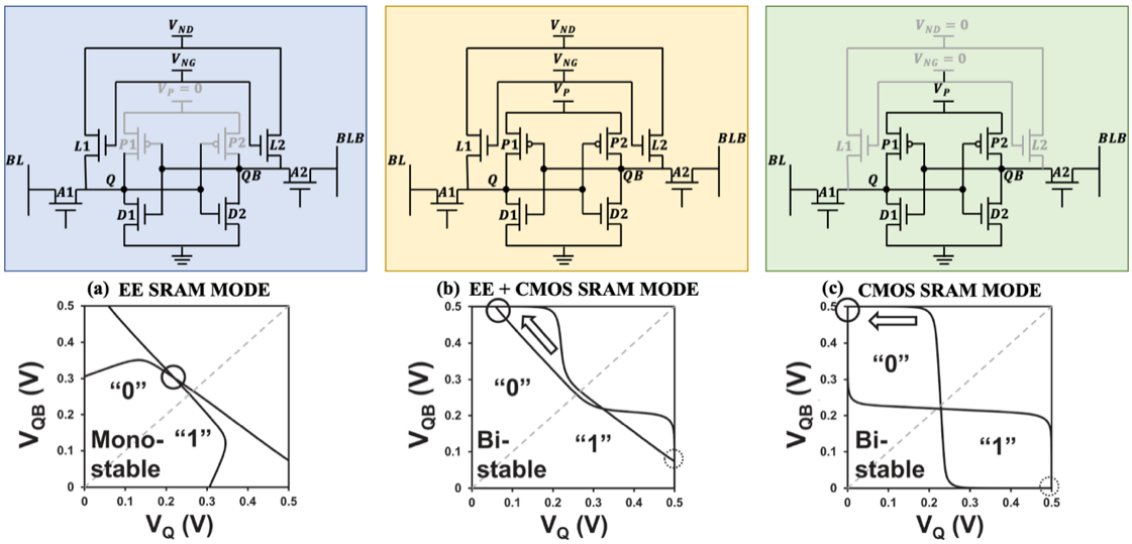


Fig. 2.8. Working mode of the 8T SRAM PUF. (a) EE SRAM for stable evaluation. (b) Transaction from EE SRAM to CMOS SRAM mode. (c) CMOS SRAM mode proposed in [30].

From this figure, the 8T bitcell is composed by a conventional 6T SRAM plus two diode-connected nMOS load transistors (i.e., $L1$ and $L2$).

During the PUF evaluation, V_P is connected to the GND, whereas V_{NG} and V_{ND} are powered up simultaneously thus forcing the circuit to work in EE SRAM mode. Here, the output data is function of the mismatch between the diode-connected nMOS load transistors (i.e., $L1$ and $L2$) and the two nMOS driver transistors (i.e., $D1$ and $D2$). The low gain of the bitcell forces the structure to work in a monostable state as showed in Fig. 2.8 (a). This monostable behavior improves the immunity to on-chip noise since there is only one possible operative state. After the two voltages are generated (i.e., Q and QB) they need to be pushed toward V_{DD} and GND for improving the stability during the read-out phase. In [29] authors increased the V_{DD} for increasing the gain enough to bias the cells into the bistable state. However, this approach suffered from high short-circuit current thus getting worse the power consumption. In [30] the authors use a data latching scheme where instead of raising V_{DD} the monostable solution is latched by turning on V_P and switching the PUF from EE SRAM to EE + CMOS SRAM intermediate state as shown in Fig. 2.8 (b). After that V_{NG} and V_{ND} are cut off the cell enters in the CMOS SRAM mode as shown in Fig. 2.8 (c). In this mode the short-circuit currents are suppressed thus reducing the required power consumption for reading the data. Moreover, V_Q and V_{QB} are pushed to full-rail voltages by the cross-coupled CMOS inverters. The design was tested in 130-nm CMOS technology. Measurement results demonstrate that this solution allows achieving the same benefits in terms of area and energy efficiency (e.g., 497 F^2 and 15.39 fJ/bit) of a conventional SRAM cell while also keeping the native instability at GK conditions (i.e., V_{DD}= 0.6 V and T = 25°C) low (e.g., 0.29% and 2.71% as BER and unstable bits, respectively) at the cost of a non-conventional SRAM cell. However, VT variations significantly affect the native stability. Indeed, this solution shows an average native BER of 1.30% and 1.37% at 0.5 V and 0.7 V, respectively, and of 2.99% and 5.76% at -40 °C and 120 °C, respectively.

2.5.2 *Delay-based PUFs*

Delay based PUFs represent the first explored solutions. These circuits translate the process variation into a measurable delay quantity then converted into a binary response. The first delay-based solution consists of N pairs of nominally identical ring oscillators (ROs) where the bit response is 1 or 0 depending on which of the two Ros in the pair is faster, e.g., if the first RO is faster (slower) than the second the output bit is 1 (0). However, the conventional RO-based architecture has poor stability because most ROs have a frequency lying around the main value of their distribution thus resulting in frequency differences close to zero. This makes the frequency comparison sensitive to on-chip noise and PVT variations. To overcome to these issues, different solutions have been proposed. One recent example is reported in [40] where the authors proposed a delay cell topology with an adjusted signal slope to amplify the impact of the transistor mismatch on the delay along with an in-situ re-calibration for improving the PUF stability. Fig. 2.9 illustrates the schematic of the solution proposed in [40] along with the simulated distribution of the frequency difference.

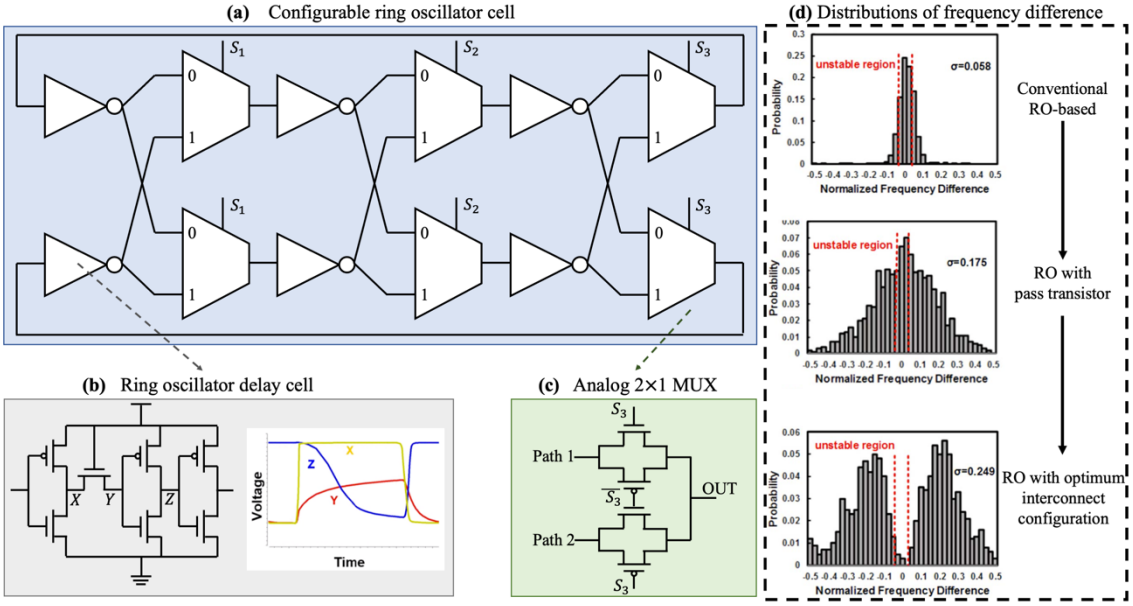


Fig. 2.9. Schematic of (a) configurable RO, (b) RO delay cell and (c) analog 2×1 MUX, proposed in [40]. (d) Simulated distributions of the frequency difference.

For increasing the spread of the frequency difference distribution, the authors proposed a different RO delay cell showed in Fig. 2.9 (b) which consists of placing a nMOS pass transistor between the inverters which is sized for being the dominant variability source in the cell. More precisely, due to the nMOS incapability of delivering the high voltage signal X, its output Y exhibits a gradual low-to-high transition because the nMOS operates in sub-threshold regime. As result, a higher delay spread can be observed at the inverter stage output Z compared to the conventional cell. This can be observed in the frequency difference distributions showed in Fig. 2.9 (d). Despite the use of pass transistor, the same figure reports that a high percentage of samples still fall in the unstable region. To further reduce the instability the selector signals can be tuned for choosing the better RO configuration during the testing time thus resulting in a significant reduction of

samples which fall in the unstable region and hence sensitive to noise and different VT conditions. The main advantage of this solution is the possibility of achieving a very low BER correctly setting the selector signals. The design was tested in 28-nm CMOS technology. Measurement results reveal that at GK conditions (i.e., V_{DD}= 0.9 V and T = 25 °C), the lowest reported BER before calibration is 1.4% which can be decreased to 0.156% with online calibration. Concerning the VT variations, the measured BER before and after calibration is, respectively, around at 8.5% and 1% at both V_{DD} corners of 1.3 V and 0.4 V and around 3% and 0.2% at -40 °C and 4.5% and 0.3% at 125 °C. However, this stability is achieved at the cost of a high bitcell area (e.g., 33,163 F^2) and required energy (e.g., 2.15 pJ/bits). Moreover, the configuration data needs to be stored in a NVM on chip. Anyway, these data do not reveal information about the PUF response.

2.5.3 Metastable-based PUFs

To this class belong PUF circuits which exploit a metastability event resolution as static entropy source for generating random bits. These circuits leverage on the natural tendency to settle to a preferred state, after entering in a metastable state, determined by mismatch after a reset or pre-charge operation. Such operative principle is similar to conventional SRAM approach and can be implemented by using cross coupled circuits for inducing a bistable behavior in the circuit. The main advantage of using metastable solutions relies on the native resiliency to power attacks. Indeed, such circuits provide both direct and negated bits thus resulting in no difference in the power profile when reading '1' or '0'. One of the most relevant solutions was proposed in [42]-[44] where the authors exploited a hybrid cross-coupled inverter circuit with a pair of pre-charge transistors for initializing the internal nodes to an unstable state. Fig. 2.10 illustrates the delay-hardened bitcell proposed in [43].

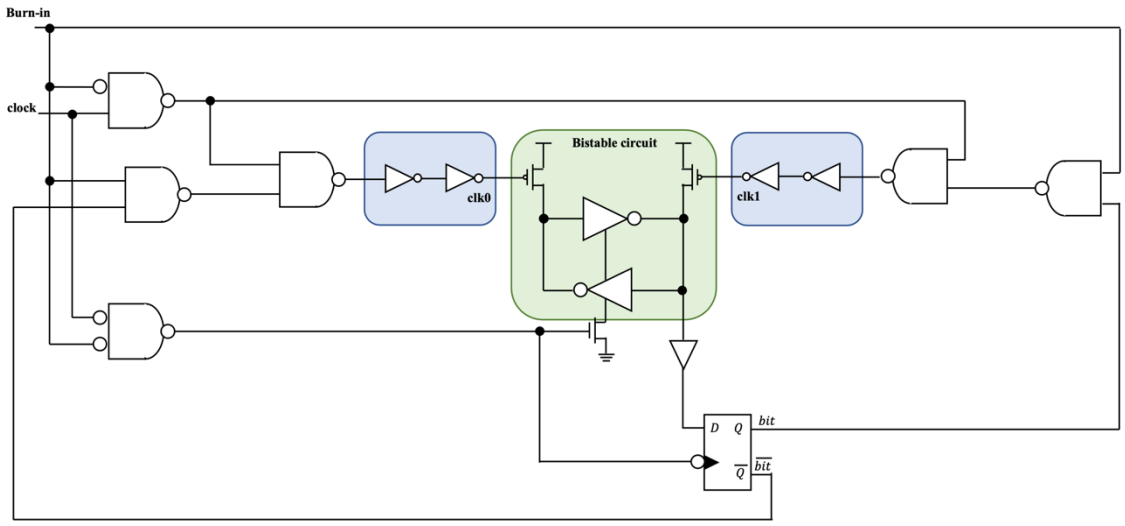


Fig. 2.10. Delay-hardened PUF circuit proposed in [43].

The bistable circuit, highlighted in green in Fig. 2.10, evaluates and collapses in one of the two possible stable states during the positive phase of the clock. Indeed, the circuit resolves the metastable state toward V_{DD} or ground according to the strength mismatch between the two minimum sized cross coupled inverters. Moreover, random variations in the peripherical circuits,

which carry the clock signal at the input of the bistable circuit, also allow increasing the uncertainty into PUF resolution dynamics. The flip-flop captures the evaluated PUF output at the negative clock edge. Since the PUF value is not directly read from the bistable circuit, the architecture also adopts clock gating to pre-charge the cross-coupled inverter to V_{DD} between the two consecutive evaluations. In this state, the two transistors in both legs of the bistable circuits receive the same gate voltage thus enabling the isoaging of the circuit (e.g., thus limiting the long-term stability degradation). However, this circuit suffers from the same issues which affect the conventional SRAM approach. Indeed, in the case of small mismatch the metastability resolution toward one of the two stable states can be determined by the thermal noise or VT variations thus resulting in unstable bits. The design was tested in 14-nm tri-gate CMOS technology. Measurement results reveal a BER at GK conditions (i.e., V_{DD}=0.65 V and T = 70°C) of 5.76%. To make this solution suitable for being implemented in hardware security applications, the circuit requires stability enhancement techniques explained in the sub-section 2.5.

2.5.4 Monostable-based PUFs

This class of PUFs refers to cells which generate a static output with only one stable state. The main differences over the previous reported class are: (*i*) The static behavior ensures that the output is independent from coupling noise (i.e., the static design provides strong resiliency to noise and fluctuation in the environmental conditions because of the absence of dynamic switching events) and insensitive to routing as opposed to the delay-based solutions; (*ii*) The monostability (i.e., the existence of only one stable state) assures that the correct output (i.e., the value associated to the stable point) is always delivered even when occasionally transient noise flips the bitcell as opposed to the metastable-based solutions where once the metastability is resolved toward one of the two stable points due to noise there is no chance of returning in the correct output without re-evaluating the PUF cell.

Typically, PUF solutions which belong to this class are composed by a transformation block which translate process variation into a voltage or current signal along with a digitizer for generating the output bit. The integration of the conversion block within the bitcell ensures that coupling effect along with a long routing do not affect the uniqueness, at the cost of higher standby power and lower area efficiency compared to the solutions which share the same digitizer with different cells. One of the most relevant works was proposed in [49]-[51] and basically consists of connecting two back-to-back regulated cascode current mirrors (RCCMs) (i.e., pMOS and nMOS RCCMs) illustrated in Fig.2.11. The basically idea was previously proposed in [49] and consisted of connecting only two cascode current mirrors (CMs) in which the mismatch between the two delivered currents was first translated into a voltage signal through a high output impedance and then converted in a binary response by an output buffer. However, this solution showed three main limitations: (*i*) reliability issues under voltage scaling; (*ii*) occasional non-full swing voltages due to small mismatch were translated into instable bits at the output of the buffer; (*iii*) temperature variations affected the pMOS/nMOS skew thus getting worse the stability. To overcome these limitations, the authors proposed a novel bitcell design in [50] where they used a RCCM instead of a classic CM along with the Muller C-element instead of the conventional buffer as illustrated in Fig. 2.11.

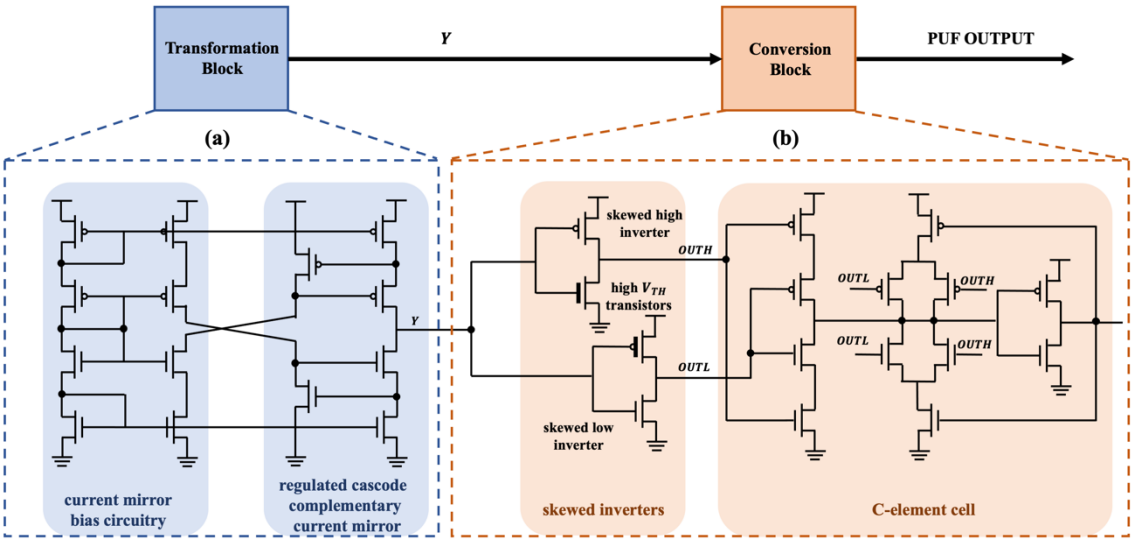


Fig. 2.11. Design concept of the bitcell proposed in [32] along with the schematic of (a) bitcell core based on RCCM and (b) conversion block based on C-element cell.

The RCCM exhibits a larger output resistance compared to the previous approach thus resulting in a higher spread even with small mismatch. Moreover, it also contributes to the overall pMOS/nMOS current mirror mismatch. As an additional advantage, the RCCM can operate at lower voltage thus improving the reliability under voltage variations. The use of the Muller C-element instead of the buffer increases the noise immunity by 110 mV thus reducing the impact of the voltage fluctuations at the Y node. Indeed, the output of the Muller C-element switches only with voltage fluctuations enough large to cross both the logic thresholds of the skewed inverters. Authors also included in the design a temperature compensation scheme for assuring that the pMOS/nMOS strength is maintained constant across temperatures by adjusting the pMOS body voltage. The design was fabricated in 40-nm CMOS technology. Measurement results at GK conditions (i.e., V_{DD}= 0.9 V and T = 25 °C) show a native percentage of unstable bits (worst BER) of 3.48% (0.81%) in 5000 evaluations. Concerning the VT variations, the circuit shows that the native unstable bits (BER) sensitivity to V_{DD} variations is 3.6%/V (7.2%/V) whereas the sensitivity to the temperature variations is 0.018%/°C (0.032%/°C). The proposed solution also shows a bitcell energy of 1.02 fJ/bit and a readout energy of 56.5 fJ/bit. This is achieved at the cost of a large bitcell area (i.e., $3644F^2$).

2.5.5 Hybrid PUFs

This class of PUFs combines two different PUF topologies with the aim of exploiting some advantage of both. Several research groups demonstrate the possibility of using delay-based solutions along with monostable-based circuits for implementing PUF instance with very low native BER. The main reason is that from one side the delay-based solutions can integrate the mismatch information over the time thus averaging the noise effect at each stage. However, the tight distribution centered on the mean delay value makes the circuit sensitive to PVT variations as well as to noisy comparison. On the other hand, monostable-based solutions exploit the only one stable working point for ensuring that the correct output is always delivered despite the occasional flips due to the noise effect on the first stage. This is because these circuits typically

implement one high-gain stage along with conversion circuitry for binarizing the response. However, monostable solutions demonstrated to be able of operating in wider voltage and temperature ranges while keeping the instability low. Recently, hybrid solutions have been proposed which allow exploiting the benefits of both delay-based and monostable-based solutions. For example, in [60] the authors propose a PUF with an amplification process mismatch in an oscillator collapse topology. The proposed PUF cell is comprised of a process-to-voltage converter (PVC) and a process-to-time converter (PTC) as illustrated in Fig. 2.12.

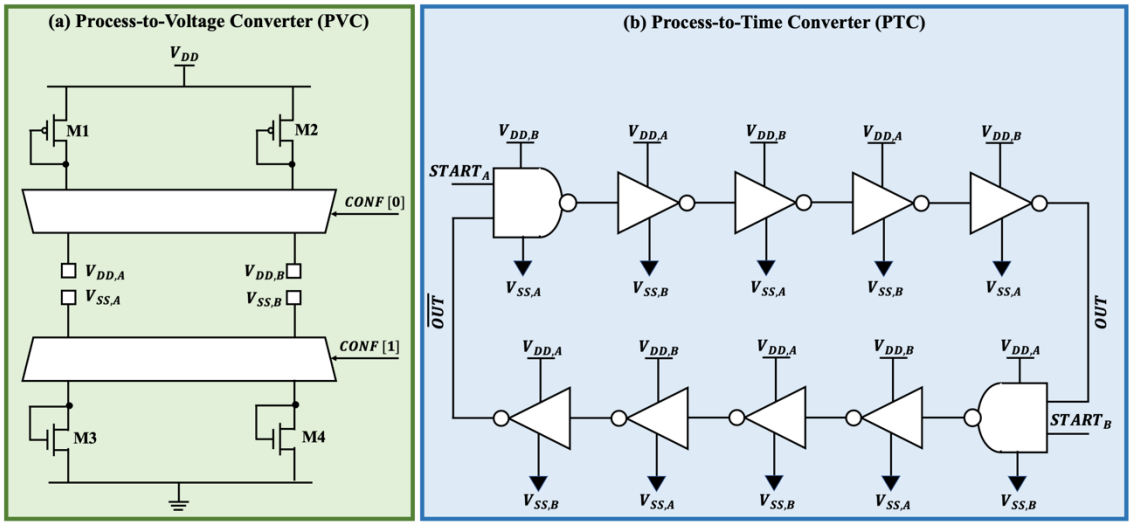


Fig. 2.12. Circuit design of the PUF proposed in [42].

The PVC compries two pull-up pMOS transistor and two pull-down nMOS transistors which provide the supply rails (i.e., $V_{DD,A} - V_{DD,B}$ and $V_{SS,A} - V_{SS,B}$ respectively) to the PTC. The latter consists of ten-stage RO which is composed by current-starved inverters (i.e., the supply voltage of the inverter is provided by the PVC in an alternate way). The race starts after the raising transition of the $START$ signals through two NAND gates. Due to the mismatch these two signals flow through two electrically different paths. In this way, one edge overtakes the other thus determining the PUF output state. The static entropy source relies on the mismatch of both PVC (i.e., variations in the supply rails) and PTC (i.e., variations in the logic gates). For this reason, the signal $CONF$ [0] and $CONF$ [1] can be employed for aligning the polarities of the effects by PTC and PVC during the testing time. Indeed, a higher V_{DD} along with a lower V_{SS} makes the inverter faster compared to the opposite combination which makes the inverter slower. These effects need to be added to the variability in the PTC during the testing time. The collapse of oscillation occurs when the total delay difference between the two paths reaches one-half of the period of both paths (i.e., RO_1 and RO_2 which represent the upper and the lower path respectively). The goal is to increase the $t_{interval}$ (i.e., the given time interval for the catch-up which is proportional to the n delay stages) for improving the robustness against noise (i.e., the noise is averaged out for a longer time) while maintaining low the cycles to collapse (CTC) and this is achieved by using ten-stage RO in the PTC along with the PVC. The design was tested in 40-nm CMOS technology. Measurement results prove the effectiveness of the proposed solution. Indeed, the native measured percentage of unstable bits (BER) at GK conditions (i.e., V_{DD}= 0.9 V and T = 25 °C) is 0.39% (0.027%). Concerning the VT variations, the circuit shows a very low sensitivity to voltage (e.g., from 0.7 V to 1.4 V) and temperature variations (e.g., from -40 °C to

120°C) when supported by stability enhancement techniques described above. This is achieved at the cost of a very large area per bit of 21,675F^2.

2.5.6 Active PUFs

This class of CMOS PUFs basically consists of solutions in which the translated static entropy source does not come from the CMOS process variations, but it is actively implemented during the testing time. The most relevant approach implements the oxide breakdown (BD) as reported in [65] and [66]. Despite these solutions achieve a near-zero BER, they also violate two important PUF requirements: (*i*) a physical inspection (e.g., delayering and imaging) can potentially reveal the secret key; (*ii*) the PUF response is available even when the chip is powered off since the information is stored in a non-volatile manner. Recently, a research group proposed an interesting solution based on the soft oxide breakdown (SBD) with a self-limiting technique [62]-[64] with the aim of achieving the same performance of the BD approach while reducing the vulnerability to the physical inspection of the device. The authors proposed a three-transistor (3T) structure which consists of two minimum-sized nMOS and one pMOS as illustrated in Fig. 2.13.

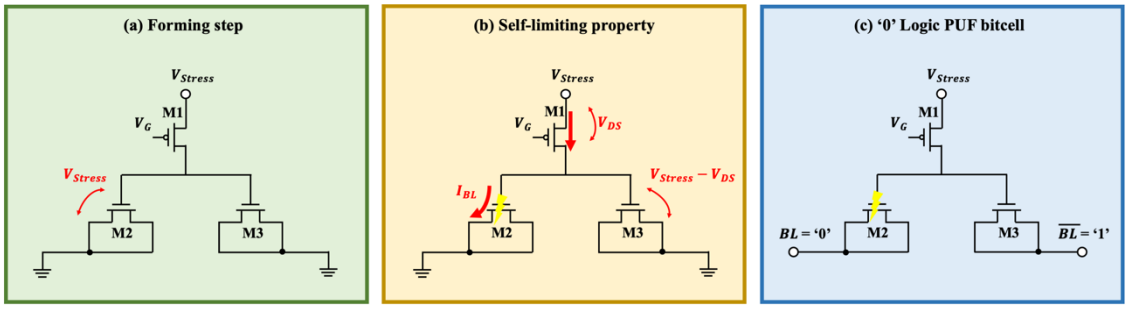


Fig. 2.13. Soft oxide BD procedure of the PUF bitcell proposed in [64]. (a) Forming step, (b) self-limiting mechanism and (c) generation of bit '0'.

During the forming step the two nMOS transistors are subjected to a high stress voltage. After the BD occurs (i.e., after the BD time,t_{BD}, which is function of the applied voltage) in one of the two transistors a current will flow in the BD spot thus inducing a voltage drop across the pMOS. This reduces the stress voltage on the health nMOS thus avoiding the generation of other BD spots. The magnitude of the observed current varies depending on the shape and size of the BD spot. The generated bit can be read by using a sense amplifier (SA) which also allows improving the resiliency to side-channel attacks. This solution was tested in 40-nm CMOS technology. Measurement results demonstrate its effectiveness. Indeed, this circuit works well as large is the difference between the BD current and the leakage current (i.e., trap-assisted tunneling, TAT) which flows in the health transistor. The measured BER is 0% for voltages above 0.9 V regardless of the temperature. For lower voltages the two currents become comparable thus resulting in a higher instability. Indeed, at 0.9 V the zero-BER condition is maintained until 120 °C. On the other hand, when decreasing the supply voltage down to 0.8 V the percentage of unstable bits (BER) increase over 0.4% (0.09%) at room temperature. A further increase can be observed when increasing the temperature. In fact, flipped bits increase above 1.5% at 60°C. These results highlight that this solution is not suitable for low-voltage operations. This circuit also shows a relative compact bitcell area of 1875$F^2$with an energy of 51.8 fJ/bit.

2.5.7 Other PUFs

The PUF solutions described above strictly refer to CMOS implementation. Nowadays, several others PUF topologies have been proposed. For example, Contact PUF [67] which exploits the randomness in the contact (i.e., the vertical interconnect layer between the metal and the silicon) forming process for generating the secret key. Indeed, during the fabrication process the contact formation probability is function of the contact size. During the design phase, if the contact size is chosen enough small its probability of being formed can be set to the 50%. This is because in the formation process, after the dielectric is deposited on the silicon substrate, a contact hole is formed by the etching process and then it is fully filled with metal (i.e., typically tungsten). If the contact size is too small an error can incur during the etching and the filling phases thus resulting in a shorted contact or open contact. This solution shows a zero-native instability but presents the same drawback of the active PUFs for which the key information is stored in a non-volatile manner thus exposing the secret key to physical inspections. Recently several research groups are focusing on the PUF implementation in FPGA systems [68]-[75] to encrypt the bitstream, for example. Typically, FPGA-based solutions implement delay-based and metastable-based PUF architectures, due to already placed physical circuits. However, these solutions present many drawbacks. The common delay-based implemented architectures are the arbiter PUF (i.e., APUF) and the ring oscillator PUF (RO-PUF). Due to physical layout restrictions, it is not simple to implement these circuits while achieving good performance in terms of uniqueness and reliability. Indeed, FPGA designers tried in the years of proposing different architectures with manual routing and/or forcing the synthesizer to route such signals in a constrained way [73]. Metastable-based solutions implemented on FPGA show the same routing issues thus requiring a large effort for implementing them. Indeed, synthesizer rules inherently break the symmetry of the cross-coupled gates (i.e., typically latches) by sharing the control signal thus making this solution difficult to implement (i.e., each latch should be implemented in a different slice). Other relevant solutions involve the use of emerging devices [76]-[88] such as MTJ, ReRAM and EGFET. In the first two cases, the authors typically set the writing phase for having a failure probability of 50%. However, the random data is generated in a non-volatile manner thus violating one of the fundamental PUF requirements. On the other hand, EGFET refers to printed electron devices whose variability is exploited for implementing a PUF in a SRAM fashion.

2.6 Strong PUF implementation

Strong PUFs refer to architecture able of generating a number of CRPs that increases exponentially with the number of physical implemented cells. Like the weak PUFs, the strong PUFs can be categorized in different classes depending on the way in which the process variations are translated into a measurable quantity and then converted into a binary response. For example, we can find delay-based PUFs [89]-[91], SRAM or SRAM-based PUFs [92]-[93] and monostable-based PUFs [94]-[96]. The abundant availability of CRPs increases the level of security if the computational effort to predict the responses with a model is large enough. However, the generated bits in most of the strong PUF architectures relies on the linear

combination of multiple sources of randomness thus making these solutions vulnerable to machine learning attacks. Accordingly, the major challenge in strong PUFs is to combine the multiple entropy sources in a non-linear manner thus increasing the reliability to external modeling attacks.

2.6.1 Delay-based PUFs

The first explored delay-based topology is the arbiter PUF [89] which consists of two nominally identical delay paths (i.e., they have the same nominal delay) along with an arbiter. An applied input signal propagates through the two paths and arrives to the arbiter in different moment due to the process variations. The arbiter provides a bit '1' or '0' at the output depending on which path wins the race (i.e., which of the two paths is faster). The arbiter is typically composed by a Set-Reset Latch (e.g., two cross-coupled gates) or by a delay flip flop (DFF). The main problems of this architecture are: (*i*) typically these paths have a delay lying around the main value of their distribution, making the these delay values close to each other thus resulting in a poor stability since the metastability resolution of the arbiter can be influenced by on-chip noise and different environmental conditions; (*ii*) This architecture also shows high sensitivity to routing for which nominally unbalanced paths affect the output randomness thus increasing the predictability of the response; (*iii*) Since the output bit is generated by linearly summing the delay of each stage the CRPs can be predicted by a combination of advanced machine learning and side-channel attacks. Recently, many research groups proposed different solutions to overcome the linearity correlation between input (challenge) and output (response) such as the challenge and/or response obfuscation (e.g., with ah hash function) or designing difference bitcell structures which combine the multiple entropy sources in a non-linear manner. For example, in [90] the authors exploit the concept of oscillation collapse in a ring with an even number of inverter gates as illustrated in Fig. 2.14.

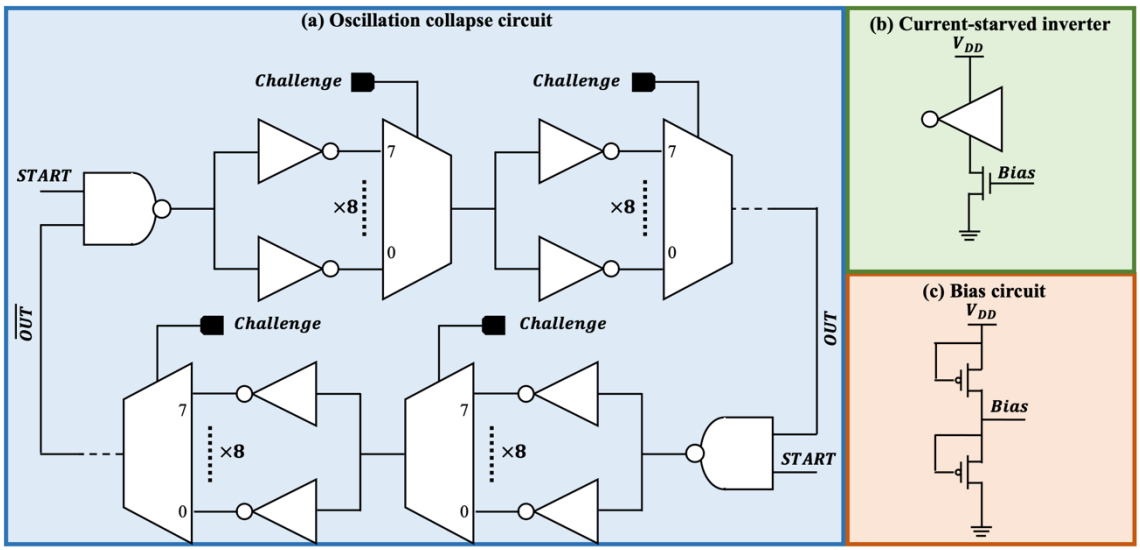


Fig. 2.14. Example of delay-based strong PUF [90] with (a) the oscillation collapse circuit, (b) the current starved inverter and (c) the bias circuit.

The circuit in this figure translates the process variations into a binary response by injecting two edges into an even-stage RO. The two edges travel in two electrically different paths, due to the mismatch, causing delay accumulation which leads one edge to overtake the other, thus collapsing the oscillation. Depending on which path is faster the output node will collapse in a logic '1' or '0'. The use of a current-starved inverter, showed in Fig. 2.14 (b), instead of the conventional structure allows increasing the response stability by ensuring that the variation of the footer nMOS dominates the total delay. The CTAT circuit showed in Fig. 2.14 (c), generates a bias voltage on chip as well as performs a first-order temperature compensation of the footer current to reduce the PUF temperature sensitivity. Finally, thresholding is used for improving the stability of the response. Indeed, discarding the CRPs with a number of cycles to collapse higher than a certain threshold helps to suppress the BER across a wide range of temperatures and voltages. The solution shows a core area of 528,125F^2 and zero-BER at GK conditions (i.e., V_{DD}= 0.9 V and T = 25 °C) when discarding around 30% of the CRPs. The use of the current starved inverter with a CTAT bias circuit along with a proper thresholding ensures a $< 10^{-8}$ BER across -25 to 125 °C temperature range and 0.7 to 1.2 V voltage range. However, this architecture does not introduce an explicit nonlinearity thus still making this solution potentially vulnerable to machine learning attacks.

2.6.2 SRAM and SRAM-based PUFs

This class of PUFs is typically used for chip ID. Recently, many groups proposed different approach for expanding the CRP space of SRAM/SRAM-based PUFs thus making them suitable for device authentication. One example is reported in [92] where the authors exploit a sequence-dependency instead of the conventional power-on state. Fig. 2.15 illustrates the array architecture along with the schematic of the single SRAM cell. Here, one bit response is generated by enabling simultaneously a pair of opposite-valued cells which share the same bitline. Basically, any two bitcells are connected and initialized with complementary data by asserting their word-lines. The output bit depends on the relative strength of all the 12 transistors of the two enabled bit-cells and their initial value.

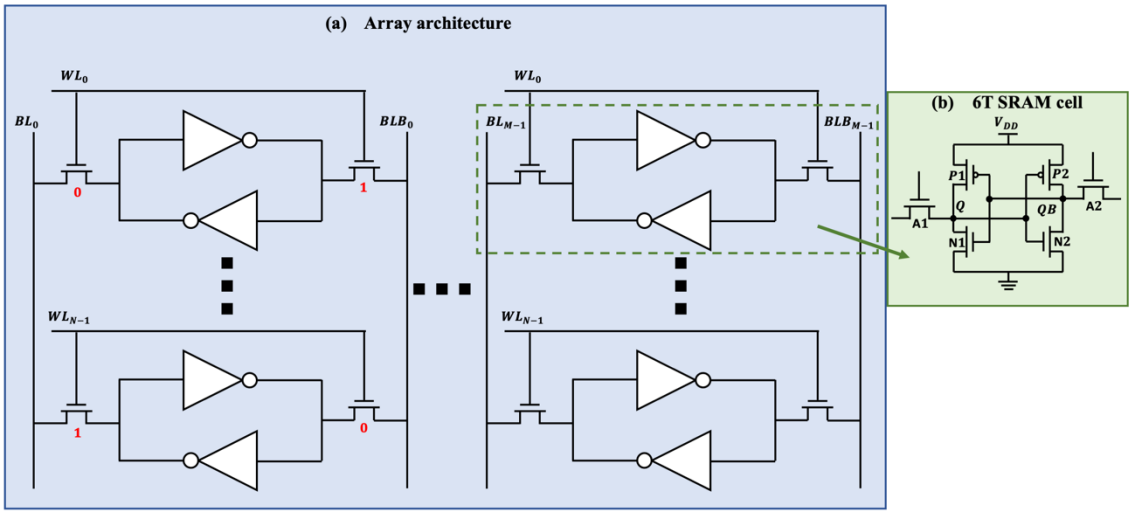


Fig. 2.15. Schematic of (a) matrix of the nonlinear sequence-dependent architecture and (b) single 6T SRAM cell of the solution proposed in [92].

This approach introduces nonlinearity thanks to the dependence of the output on the sequence of enabled pairs. This solution shows an area per bit of 388 F^2 along with a native BER at GK conditions (i.e., V_{DD}= 0.7 V and T = 25) of 3.17%. However, this approach suffers from high sensitivity to VT variations. Indeed, the BER increases up to ~8% when increasing (decreasing) the V_{DD} up (down) to 900 (550) mV and up to ~10% when increasing the temperature up to 80 °C.

2.6.3 Monostable-based PUFs

Monostable-based solutions can be also used for designing strong PUF architectures with strong nonlinearity. For example, in [96] the authors exploit a sub-V_{TH} current array for comparing the current delivered by two transistor arrays in deep sub-threshold. Fig. 2.16 illustrates the schematic of the SCA PUF along with that of the single sub-threshold current array (SCA). The output bit is generated by comparing the output voltages of the two SCAs. Each SCA is composed by $n \times k$ unit cells. Each unit cell consists of two transistors, i.e., M_{ijx} and M_{ij}. The latter refers to the stochastic transistor with minimum sizing for optimizing the V_{TH} variability. On the other hand, the parallel transistor M_{ijx} is optimized for reducing its variability. The bias of each array is chosen so that the currents in each branch keep the diode-connected stochastic transistors in sub-threshold region regardless of the input control signals. The sub-threshold operation is ensured via current sources M_{c1} and M_{c2}. Indeed, when the control signal (i.e., the challenge) C_{ij}= 0 the stochastic transistor is shorted by the switch transistor (i.e., M_{ijx}). On the other hand, when C_{ij}= 1 the stochastic transistor (i.e., M_{ij}) operates in sub-threshold region and contributes to the output voltage. The two current sources are nominally identical and biased by the same voltage. This implies that in absence of mismatch the two output voltages are equal. However, the V_{TH} randomness leads to have differences in the output voltages.

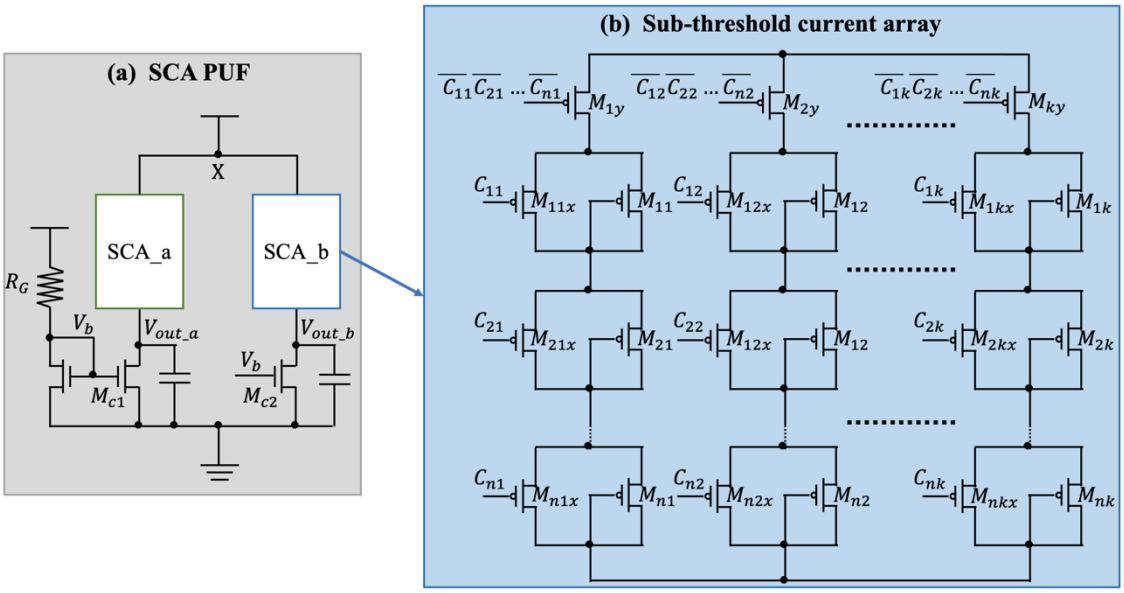


Fig. 2.16. Schematic of (a) SCA PUF and (b) sub-threshold current array proposed in [96].

These differences are then digitized by the comparator. The nonlinearity behavior between the entropy sources and outputs arises from the exponential relationship between the current and the

V_{TH} of each transistor. This behavior ensures a good resilience against modeling attacks. Moreover, this solution shows an area of $47{,}929F^2$ and a worst native BER of 9% which can be decreased down to 0.4% by discarding the 42% of the CRPs. Regarding the VT variations, the worst measured BER increases up to ~4% when increasing (decreasing) the V_{DD} up (down) to 1.32 (1.08) V and up to ~12% when increasing the temperature up to 80 °C.

2.7 Stability enhancement techniques

Typically, the raw stability of a PUF instance is inadequate for ensuring a quasi-zero error probability during a high fraction of the device life. Indeed, additional techniques are always required for reducing the native instability enough to achieve the targeted KER. Fig. 2.17 illustrates an example of how reducing the PUF instability during the chip lifetime.

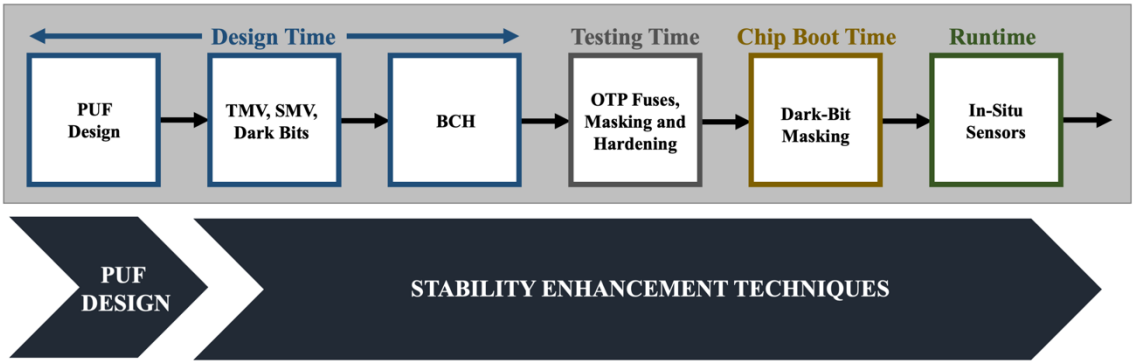


Fig. 2.17. Examples of how reducing the PUF instability during the chip lifetime.

This figure points out that several techniques can be adopted at different time during the design phase and/or during the chip lifetime, ranging from the testing time (i.e., before providing the chip to the consumer) to the normal operations. In this sub-section some of the most used techniques will be described.

2.7.1 Techniques at design time

PUF stability can be improved at design phase by optimizing the circuit and the physical design of the bitcell for reducing the probability of having an unstable response as much as possible. For example, in [50] the authors used a C-Muller element instead of a conventional buffer for increasing the noisy immunity around 100 mV. To further improve the stability at design time techniques such as temporal majority voting (TMV) and spatial majority voting (SMV) can be adopted. These techniques exploit a redundancy (i.e., temporal and spatial redundancy for TMV and SMV respectively) for reducing the PUF instability. In particular, TMV relies on evaluating the same instance for a given challenge for an odd number of times so that the considered output will be the most frequent one in the collected responses. This technique helps reduce occasionally flips induced by transient failure mechanisms such as on-chip noise. Indeed, several works exploit the time redundancy for achieving a zero-native instability at GK conditions. However, it cannot correct error caused by permanent physical failure mechanisms such as aging or variations in the environmental conditions (i.e., temperature and voltage variations). On the other hand, SMV

relies on combining multiple PUF bits for a given challenge to generate a single response bit. Obviously, spatial redundancy is more effective for mitigating errors caused by variations in the environmental conditions since the bitcells with higher sensitivity to voltage and temperature variations are statistical infrequent. For example, in [40] the authors include extra circuitry in the design for making the bitcell configurable by combining different delay elements. In this way, the infrequent combinations at which correspond a higher voltage and temperature sensitivity can be canceled out. ECCs can be also used for correcting a fixed number of bits. This solution can effectively treat flips in the PUF responses due to on-chip noise or different environmental conditions. However, this approach is very expensive in terms of area and energy. Indeed, the ECC area and energy costs are typically two orders of magnitude larger than the PUF itself and they increase linearly with the number of correcting bits. The most used ECCs are Bose-Chaudhuri-Hocquenghem (BCH) and the Fuzzy extractor.

2.7.2 Techniques at testing time

During the testing time several techniques are used for marking and reducing the unstable bits. One of the most used consists of masking unstable bits which cannot be corrected by the ECC so that they can be cancelled out or stabilized with other techniques. For example, in [43] the authors exploit the TMV technique for enabling the soft dark bits masking. Indeed, each bit which results unstable under repeated evaluations is marked as unstable. However, this approach allows achieving information only about bits sensitive to the on-chip noise. In [29] the authors use the V_{SS} bias to detect hidden dark bits by shifting the mismatch distribution positively or negatively. In this way, the bits that appear unstable under repeated evaluations are marked us unstable. The main difference over [43] is that the shift modulation ensures of achieving information about the marginally stable cells (i.e., cells that could appear stable at GK conditions but unstable with a slight variation in the environmental conditions). Hardening is another effective technique which allows improving the native stability by locally (i.e., applying this technique only on unstable/marginally stable cells) aging the PUF array. However, this approach is not feasible in IoT nodes in view of their tight cost constraint. In [43] the authors employ the burn-in technique for exposing the PUF array to elevated supply voltages and temperatures for some hours so that the negative/positive bias temperature instability (NBTI/PBTI) aging reinforces preexisting biases and improves the cell stability. In [30] the authors exploit the hot carrier injection (HCI) burn-in for reinforcing the cell stability. They apply a high drain voltage (e.g., 3.3 V) along with an adequate gate voltage (e.g., 1.0 V) to bias the transistor in saturation region. In this way, the high electric field at the drain side gives to the carrier enough kinetic energy so that they get injected into the gate oxide. This generates interface states and results in a positive V_{TH} shift thus reinforcing the marginally stable cells.

2.7.3 Techniques at boot time

Dark bit masking is often used at chip bot time for marking the cells which remain unstable even after the previous techniques. This technique usually considers the cells that are unstable at a determined conditions due to on-chip noise without considering the impact of voltage and

temperature variations. Indeed, cells which are unstable at GK conditions (i.e., whose output flips under repeated evaluations) surely show an unstable behavior when varying the environmental conditions.

2.7.4 Techniques at runtime

Recently some techniques are used during the normal chip operations for mitigating the effect of variations in the environmental conditions. Indeed, the use of on-chip sensors allows reducing the percentage of discarded bits as well as potentially estimating the number of necessary correction bits in the ECC. The first benefit refers to the possibility of performing actions as consequence of variations in the environmental conditions. For example, in [50] the authors embed an imbalance sensor which allows of actively tuning the pMOS body voltage for balancing the pMOS/nMOS strength ratio under temperature variations. The same authors in [51] combine the data from the instability sensor with a BER machine learning model for predicting the number of necessary correction bits. In this way, the ECC can operate with the instantaneously bit instability instead of the pessimistic one (i.e., the worst-case scenario) extracted by PVT testing. This allows improving the energy efficiency of the ECC.

2.8 Conclusion

This chapter provided an overview of the PUF research field mainly focusing on silicon PUF solutions. These solutions exploit random process variations in CMOS manufacturing processes or active random techniques at testing time, such as oxide breakdown, for generating a unique chip ID. PUFs represent emerging cryptographic primitives which could find application in many important cryptographic protocols such as low-cost secure authentication, cryptographic key generation, remote secure sensing, and so on. However, for being suitable for these applications these solutions need to satisfy some important metrics like uniqueness, randomness, reproducibility, physical unclonability, etc. In the last few years, several PUF solutions have been proposed with very interesting properties. Unfortunately, designing a PUF circuit while meeting all the required features is still a challenge. Indeed, for PUF solutions based on transistor mismatch the main challenge is suppressing the native instability (without using stability enhancement techniques) under PVT variations. Among the different proposed solutions hybrid and active PUFs achieve a near-zero BER across a wide range of voltage and temperature variations. However, the former requires a very large bitcell footprint whereas active PUFs store the PUF information in a non-volatile manner.

Chapter 3
Voltage Divider Based CMOS PUF

The main contents are extracted from

- **M. Vatalaro,** R. De Rose, M. Lanuzza, and F. Crupi, "Static CMOS Physically Unclonable Functions Based on 4T Voltage Divider With 0.6%−1.5% Bit Instability at 0.4−1.8 V Operation in 180 nm," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 8, pp. 2509−2520, 2022.
- **M. Vatalaro,** R. De Rose, M. Lanuzza, and F. Crupi, "Stability-Area Trade-off in Static CMOS PUF Based on 4T Subthreshold Voltage Divider," 29th IEEE International Conference on Electronics Circuits and Systems (ICECS 2022), 2022.

3.1 Introduction

The proposed weak PUF solution consists of a voltage divider between two nominally identical sub-circuits (i.e., top circuit, TC, and bottom circuit, BC) as core block (i.e., the block which translates the process variations into a measurable quantity) along with a conversion block for digitizing the output, as shown in Fig. 3.1. Nominally, the two sub-circuits in the core block conduct the same current and the voltage at X node (i.e., V_X) is equal to $V_{DD}/2$ thanks to the core symmetry (i.e., the equivalent resistances of the two sub-circuit are nominally the same, $R_{TC} \equiv R_{BC}$). However, random mismatch between TC and BC causes these currents to be different. The large impedance at X node (i.e., R_X), provided by the conversion block, translates such current differences into a large voltage deviation. The V_X voltage is then converted into a binary response by the conversion block. Embedding the conversion block in the bitcell also ensures higher reliability and better uniqueness compared to solutions in which the comparator is shared within the column or the whole array [47]. Indeed, the comparator must be designed with high gain and accurate offset cancellation for avoiding systematic bias in the response thus degrading the randomness as well as the uniqueness. This requires high effort in terms of power and complexity and even when the offset is properly cancelled out these shared comparators suffer from long wires and coupling effects. The proposed solution belongs to the class of static monostable PUFs. The fully static design along with the monostable behavior avoid random noise associated with dynamic transients and ensure that the correct output is always delivered even when noise occasionally flips the output bit. Moreover, the adoption of a voltage divider between two

nominally identical sub-circuits (i.e., TC ≡ BC) ensures that the randomness is always guaranteed even at different PVT corners as well as an inherent resilience to VT variations since the two sub-circuits nominally have the same temperature and voltage sensitivity. This is achieved at the cost of a native larger bitcell footprint.

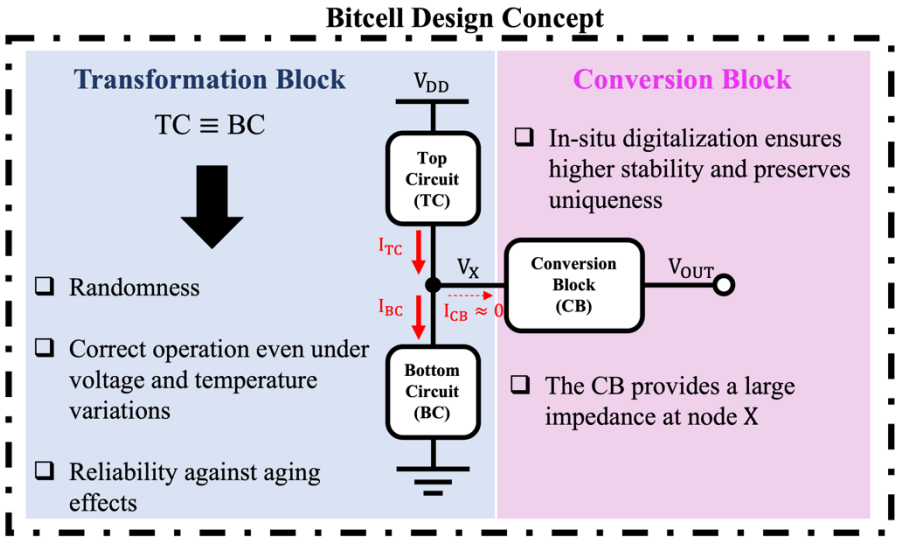


Fig. 3.7. Design concept of the proposed static monostable PUF bitcell.

The solutions explored during my PhD exploit the basic operative principle explained above while adopting different circuital solutions for implementing the sub-circuits. These different circuital solutions aim to achieve a more and more stable PUF for reducing the area and energy overhead associated to circuits implemented for suppressing the key error rate without paying an excessive cost in terms of energy and area required for generating the output bit. The following solutions have been simulated and fabricated in 180-nm CMOS technology.

3.1.1 Chapter organization

The chapter is organized as follow. Section 3.2 describes the 2T sub-threshold voltage divider. Section 3.3 analyzes the benefits and drawbacks of using a 4T sub-threshold voltage divider. Section 3.4 illustrates a possible stability – area tradeoff in the 4T solution. Section 3.5 shows the potential stability improvement when moving toward more stacked solutions. Finally, Section 3.6 concludes this chapter summarizing the achieved results.

3.2 2T Sub-threshold Voltage Divider

3.2.1 Operative principle of the 2T voltage divider

The basic idea was previously introduced in [97], where a two-transistor (2T) sub-threshold voltage divider based on identical series-connected NMOS devices with zero gate-source voltage (V_{GS}) was used. Here will be analyzed the PMOS version instead of the original NMOS version for avoiding the use of deep-n-well transistors. Indeed, for being nominally identical the body

terminal of each NMOS should be connected to the respective source terminal thus requiring the use of deep-n-well transistors. However, the parasitic currents associated to the two reverse biased diodes unbalance the voltage divider thus violating the first condition of having two nominally identical sub-circuits. Another difference over [97] is related to the way in which the bit is generated. Indeed, in [97] the output bit is generated by comparing two different cells, here the conversion block is embedded in the bitcell so that one output bit is function of only one cell. Fig. 3.2(a) and (b) illustrate the circuit of the bitcell along with the operative principle. As explained before, the bitcell consists of the 2T-core block along with an output inverter for generating the bit.

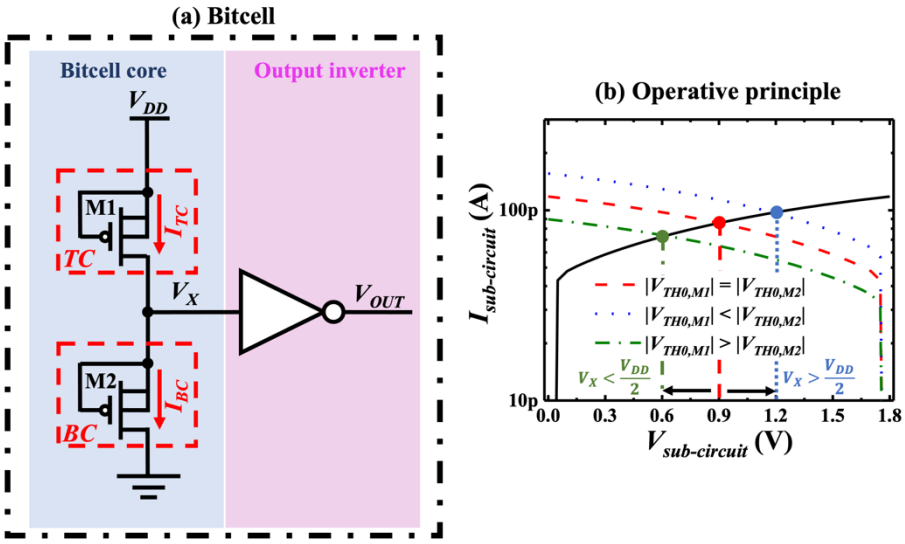


Fig. 3.8. (a) Schematic of the bitcell based on the 2T voltage divider along with (b) the operative principle.

Concerning the core circuit, it consists of a sub-threshold voltage divider between two (2T) nominally identical series connected zero-V_{SG} PMOS. Nominally (i.e., in absence of mismatch), the two transistors have the same electrical behavior and the V_X voltage is equal to the mid-supply point ($V_{DD}/2$). Due to the mismatch, the strength difference between M1 and M2 is translated into a V_X deviation from the nominal value. Indeed, if M1 (M2) is stronger than M2 (M1) the voltage drop across M2 (M1) is higher than that of M1 (M2) thus pushing V_X toward V_{DD} (ground). This principle is well illustrated in Fig. 3.2(b) which also highlights the key role played by the DIBL (i.e., drain induced barrier lowering) effect. Indeed, for a given mismatch the V_X voltage is pushed as far from the mid-supply point (i.e., $V_{DD}/2$) as low is the DIBL effect of the two transistors.

3.2.2 Design guidelines of the 2T-core

Ideally, we would have high variability in terms of M1-M2 mismatch as well as low DIBL effect for making the V_X well readable even when the mismatch is small. This highlights the trade-off between using long channel for reducing the DIBL effect and short channel for increasing the transistor variability. This trade-off is strongly technology dependent thus resulting in a different optimal sizing when scaling the technological nodes. Fig. 3.3(a)-(c) report the V_X variability (i.e., the ratio between the standard deviation, σ, and the mean value, μ) trend as function of the transistor sizing for 180-nm CMOS technology exploiting medium-V_{TH} (i.e., MVT) devices from

1k-run Monte Carlo simulations at TT corner and nominal conditions (i.e., golden key, GK, conditions) of V_{DD}= 1.8 V and T = 25°C.

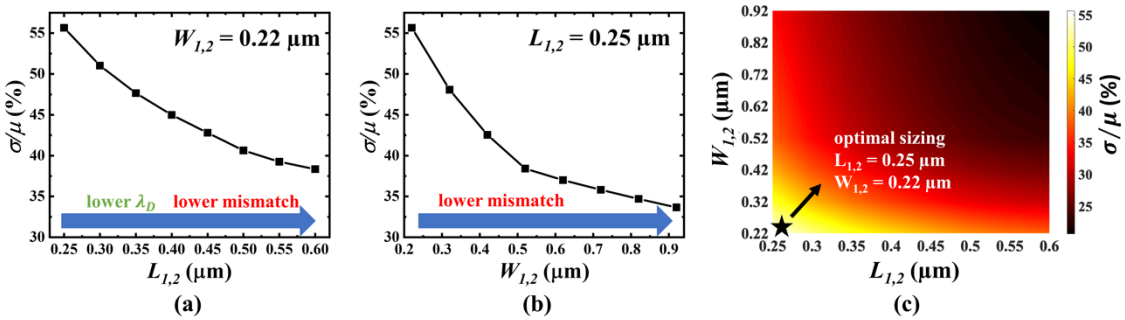


Fig. 3.3. Vx variability as function of (a) $L_{1,2}$ with $W_{1,2}$ = 0.22 µm, (b) $W_{1,2}$ with $L_{1,2}$ = 0.25µm, and (c) both $L_{1,2}$ and $W_{1,2}$ at nominal conditions of V_{DD} = 1.8 V and T = 25 °C from 1k-run Monte Carlo simulations.

Fig. 3.3(a) and (b) report the V_X variability as function of the channel length, with $W_{1,2}$= 0.22 µm, and the channel width, with $L_{1,2}$= 0.25 µm, respectively, while the map reported in Fig. 3.3(c) provides the V_X variability trend as function of both $L_{1,2}$and $W_{1,2}$. From this figure we can observe that the V_X variability decreases when increasing both $L_{1,2}$ and $W_{1,2}$. More precisely, from Fig. 3.3 (a) and (b) the increase of the channel width ($W_{1,2}$) affects more the V_X variability with respect to the increase of the channel length ($L_{1,2}$). This can be ascribed to the fact that an increase of $L_{1,2}$ results in a decrease of both DIBL effect ($\lambda_{D1,2}$) and M1-M2 mismatch ($V_{TH0,1} - V_{TH0,2}$), in this technology the second effect is dominant, whereas an increase of $W_{1,2}$ results only in a decrease of $V_{TH0,1} - V_{TH0,2}$. As result, in the considered technology the optimal sizing corresponds to the minimum one (i.e., $L_{1,2}$ = 0.25 µm and $W_{1,2}$ = 0.22 µm). Fig. 3.4(a)-(c) illustrate the statistical distributions of the electrical parameters for a PMOS MVT device with minimum sizing (i.e., L = 0.25 µm and W = 0.22 µm) from 5k-run Monte Carlo simulations. Fig. 3.4(a) reports the DIBL coefficient (λ_D) distribution from which the mean value is 17.1 mV/V with a standard deviation of 30.3 µV/V thus resulting in overall variability of 0.18%. Fig. 3.4(b) shows the threshold voltage distribution (V_{TH0}), extracted at T = 25 °C and V_{SG}= V_{SD}= V_{SB}= 0, where the mean value is -327 mV with a standard deviation of 14 mV thus resulting in a variability of 4.28%. Finally, Fig.3.4(c) illustrates the threshold voltage temperature coefficient (k_T) which shows a mean value of 881.2 µV/K along with a standard deviation of 3.6 µV/K thus resulting in a variability of 0.41%.

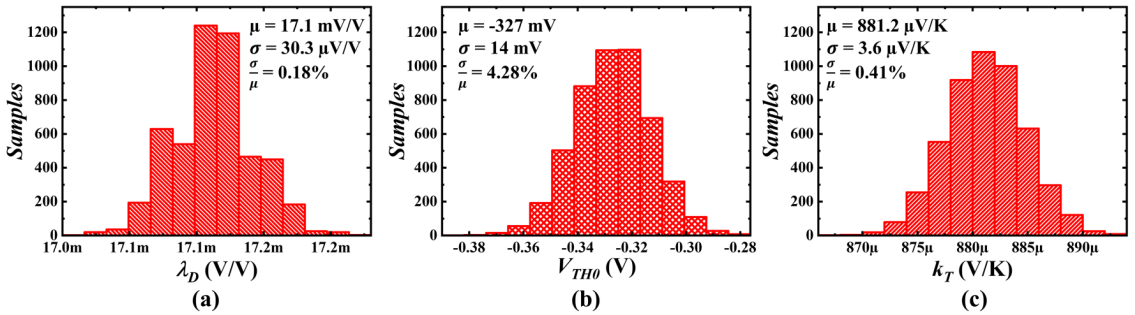


Fig. 3.4. Statistical distribution of (a) DIBL coefficient (λ_D), (b) Threshold voltage (V_{TH0}), and (c) threshold voltage temperature coefficient (k_T) from 5k-run Monte Carlo simulations for a PMOS MVT device with nominal sizing (i.e., L = 0.25 µm and W = 0.22 µm).

From a quantitative point of view, given M1-M2 working in the sub-threshold region their current can be expressed by

$$I = I_0 \frac{W}{L} \exp\left(\frac{V_{SG} + V_{TH}}{nV_T}\right)\left[1 - exp\left(\frac{-V_{SD}}{V_T}\right)\right] \quad (3.1)$$

where I_0 is the intrinsic sub-threshold current (i.e., $\mu C_{OX}(n-1)V_T^2$), W and L are, respectively, the channel width and length, n is the slope factor, V_T is the thermal voltage (i.e., $V_T = kT/q$ where k is the Boltzmann's constant, T is the absolute temperature, and q is the electron charge), μ is the carrier mobility, $C_{OX}$is the oxide capacitance for unit area, and V_{TH} is the threshold voltage. The latter can be expressed as follow $V_{TH} = V_{TH0} + \lambda_D V_{SD} + k_T(T - T_{room})$ where V_{TH0} is the zero-bias V_{TH} at room temperature (i.e., T_{room} = 25°C) and λ_D is the DIBL coefficient from the usual linear relationship $V_{TH} = V_{TH0} + \lambda_D V_{SD}$. The term in the square bracket can be neglected for $V_{SD} > 3 - 4V_T$. From (3.1) and the Fig. 3.2, the currents of M1-M2 are given by

$$I_{M1} = I_{0,1} \frac{W_1}{L_1} \exp\left(\frac{V_{TH0,1} + \lambda_{D1}(V_{DD} - V_X) + k_{T1}(T - T_{room})}{n_1 V_T}\right) \quad (3.2)$$

$$I_{M2} = I_{0,2} \frac{W_2}{L_2} \exp\left(\frac{V_{TH0,2} + \lambda_{D2} V_X + k_{T2}(T - T_{room})}{n_2 V_T}\right) \quad (3.3)$$

by equating (3.2) and (3.3) the V_X voltage can be expressed as follow

$$V_X = \frac{n_2\lambda_{D1}}{n_2\lambda_{D1} + n_1\lambda_{D2}} V_{DD} + \frac{n_2 V_{TH0,1} - n_1 V_{TH0,2}}{n_2\lambda_{D1} + n_1\lambda_{D2}} + \frac{(n_2 k_{T1} - n_1 k_{T2})(T - T_{room})}{n_2\lambda_{D1} + n_1\lambda_{D2}} + \frac{n_1 n_2 V_T}{n_2\lambda_{D1} + n_1\lambda_{D2}} \ln\left(\frac{I_{0,1}}{I_{0,2}} \frac{W_1 L_2}{W_2 L_1}\right) \quad (3.4)$$

for sake of simplicity, since M1 and M2 are nominally identical we can assume that: (*i*) $n_1 = n_2 = n_{1,2}$ and (*ii*) $\lambda_{D,1} = \lambda_{D,2} = \lambda_{D1,2}$. The latter assumption can be valid only for very low V_{DD} where the DIBL coefficient mismatch is negligible. Anyway, with these assumptions the equation (3.4) can be written as follow.

$$V_X = \frac{V_{DD}}{2} + \frac{1}{2\lambda_{D1,2}}\left(V_{TH0,1} - V_{TH0,2}\right) + \frac{(k_{T1} - k_{T2})(T - T_{room})}{2\lambda_{D1,2}} + \frac{n_{1,2} V_T}{2\lambda_{D1,2}} \ln\left(\frac{I_{0,1}}{I_{0,2}} \frac{W_1 L_2}{W_2 L_1}\right) \quad (3.5)$$

From equation (3.5) in absence of mismatch $V_{TH0,1} = V_{TH0,2}$ and $V_X = V_{DD}/2$. However, due to the mismatch, differences in $V_{TH0,1} - V_{TH0,2}$ are translated into V_X deviation from the mid-supply point as far as low is the DIBL coefficient (i.e., $\lambda_{D1,2}$). The second part of equation (3.5) refers to the temperature sensitivity. Indeed, mismatch in the k_T terms can potentially lead to a bit flip (i.e., inversion of the mismatch polarity, $V_{TH0,1} - V_{TH0,2}$) at a certain temperature. Assuming V_{TH0} variations of M1 and M2 as statistically independent in (3.5) and neglecting the logarithm term, the standard deviation of V_X (i.e., σ_{V_X}) is given by

$$\sigma_{V_X} \approx \frac{\sigma_{V_{TH01,2}}}{\sqrt{2}\lambda_{D1,2}} \quad (3.6)$$

where $\sigma_{V_{TH01,2}}$ is the V_{TH0} standard deviation of M1 and M2. To better understand how the M1-M2 mismatch is translated into a V_X deviation from the mid-supply point Fig. 3.5(a) and (b) illustrate the V_X voltages as function of the M1-M2 mismatch in terms of $V_{TH0,1} - V_{TH0,2}$ and the

2T-core gain distribution over the M1-M2 mismatch from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T = 25°C).

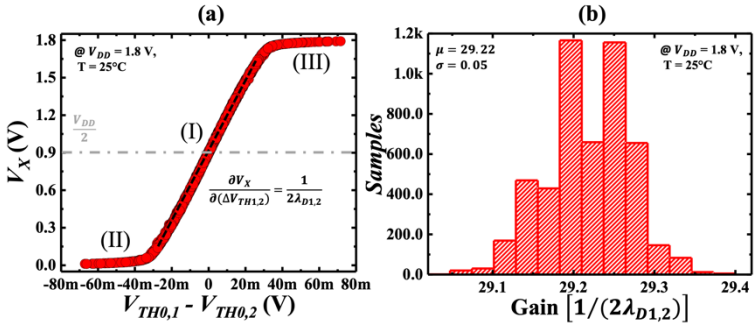


Fig. 3.5. V_X Voltage as function of the M1-M2 mismatch at GK conditions (i.e., V_{DD} = 1.8 V and T = 25°C) from 5k-run Monte Carlo simulations.

The analytical treatment developed above well describes the region (I) of Fig.3.5 in which the assumption made above (i.e., the voltage drops across the two transistors are large enough for making the terms in square bracket of (3.1) negligible) is valid. Indeed, considering a DIBL coefficient (i.e., λ_D) of 0.017 V/V, as reported in Fig. 3.4, the overall gain (i.e., $1/2\lambda_D$) is around 30. In region (II) and (III) the M1-M2 mismatch is large enough for letting the strongest transistor of the voltage divider works with a V_{SD} lower than $3-4$ thermal voltages thus approximately saturating the V_X voltages towards V_{DD} or ground according to the M1-M2 mismatch. Indeed, the slope of the characteristic in (II) and (III) is not zero but it is much lower than that in region (I). This is because the additional V_{SD} dependence of the exponential term in the square bracket leads to large current variations with small voltage (i.e., V_{SD}) variations. In this way, a lower voltage variation is required on the strongest transistor for compensating the strength difference due to the mismatch.

3.2.3 Simulations and measurements of the 2T-core

To better understand the circuit behavior under VT variations Fig. 3.6 and Fig. 3.7 illustrates the simulated and measured V_X and I_{2T} (i.e., the absorbed current by the core circuit) trends under voltage and temperature variations, respectively. Measurements of the 2T bitcell core were performed at wafer level across 20 samples by using a Cascade SUMMIT 11861B probe equipped with a Temptronic chuck temperature controller and a Keithley 4200-SCS parameter analyzer. Fig. 3.6(a) and (b) show the simulated and measured trend, respectively, of the V_X normalized to V_{DD} across voltages at T = 25 °C from 250 samples and 20 dice. Fig. 3.6(a) demonstrates that the amplification effect provided by the circuit is not enough large for pushing all the V_X voltages far from the mid-supply point when the M1-M2 mismatch is small. This is partially confirmed by the measurements in Fig. 3.6(b) which also highlights that the DIBL coefficient mismatch is not negligible especially at high voltages thus resulting in some bitflips under voltage variations. Fig. 3.6(c) and (d) show the trend of the absorbed current (i.e., I_{2T}) simulated and measured across voltages at T = 25 °C from 250 samples and 20 dice respectively. From Fig. 3.6(c) it is notable the linear relationship between the absorbed current and the supply voltage for voltages above $3-4$ thermal voltages thus indicating a relatively high DIBL effect showed by the two transistors. This trend is confirmed by the measurements as shown in Fig. 3.6(d).

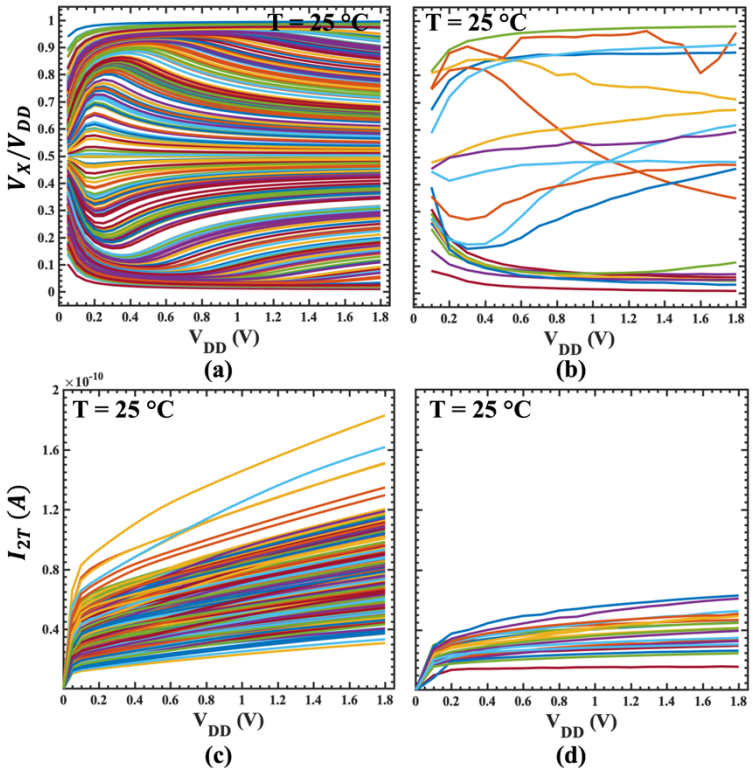


Fig. 3.6. V_X voltage of the 2T-core normalized to V_{DD} (a) simulated across voltages at T = 25°C from 250 samples and bc) measured across voltages at T = 25°C from 20 samples. Absorbed current from the 2T core (i.e., I_{2T}) (c) simulated across voltages at T = 25°C from 250 samples and (d) measured across voltages at T = 25°C from 20 dice.

Moreover, from Fig. 3.6(d) the supply current (dissipated power) of the 2T-core averaged across 20 samples decreases from 40.68 pA (87.62 pW) down to 29.73 pA (11.89 pW) when decreasing the V_{DD} from 1.8 V down to 0.4 V. On the other hand, Fig. 3.7(a) and (b) show the simulated and measured trends, respectively, of the V_X normalized to V_{DD} across temperatures at V_{DD}= 1.8 V from 250 samples and 20 dice. From Fig. 3.7(a) the V_X voltages show a nearly temperature-independent trend. However, this is not confirmed by measurements which show some bitflips. This indicates that the k_T mismatch between M1 and M2 is not negligible (i.e., the temperature part of the equation (3.5)) thus resulting in some bit flip under temperature variations. Fig. 3.7(c) shows an exponential relationship between the absorbed current and the temperature variations. This trend is confirmed by the measurements reported in Fig. 3.7 (d). Moreover, from Fig. 3.7(d) the supply current (dissipated power) of the 2T-core averaged across 20 samples increases from 40.68 pA (87.62 pW) up to 834.79 pA (1.50 nW) when increasing the temperature from 25 °C up to 100 °C. Fig. 3.6 and Fig. 3.7 point out that simulations significantly underestimate the chances of flipping the polarity of the V_{TH} mismatch between M1 and M2 and hence the instability of the PUF response under temperature and voltage variations compared to measurement results [98]. This likely arises from an underestimation of the k_T and λ_D mismatch for transistors belonging to the same schematic whose temperature and voltage dependences are somehow correlated, and from statistical correlations between mismatches in terms of V_{TH0} and k_T and of V_{TH0} and λ_D. For better explaining this problem, Fig. 3.8 summarized the statistical behavior of PMOS MVT devices with minimum sizing (i.e., L = 0.25 µm and W = 0.22 µm) under voltage and temperature variations from 5k-run Monte Carlo simulations under local variations in the considered 180-nm CMOS technology.

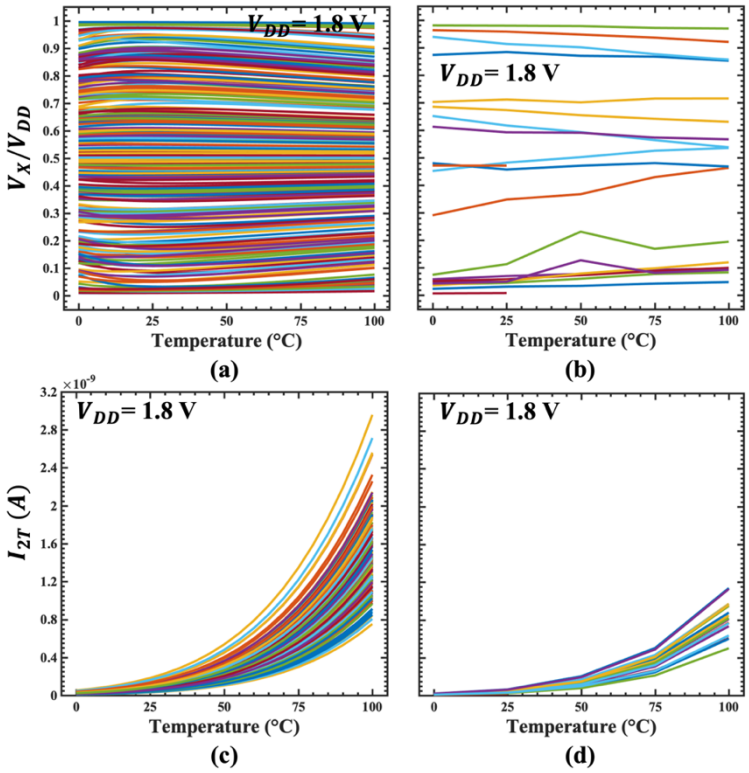


Fig. 3.7. V_X voltage of the 2T-core normalized to V_{DD} (a) simulated across temperatures at V_{DD}= 1.8 V from 250 samples and (b) measured across temperatures at V_{DD}= 1.8 V from 20 samples. Absorbed current from the 2T core (i.e., I_2T) (c) simulated across temperatures at V_{DD}= 1.8 V from 250 samples and (d) measured across temperatures at V_{DD}= 1.8 V from 20 dice.

Fig. 3.8(a) and (c) show the statistical distributions of the differences of DIBL coefficient difference ($\Delta\lambda_{D1,2}$), threshold voltage (i.e., ΔV_{TH0}), and V_{TH} temperature coefficient (Δk_T), respectively, for the two transistors belonging to the same schematic. From these figures the standard deviations are 43.2 µV/V, 19.9 mV, and 5.1 µV/K, respectively. Indeed, considering the 3σ worst-case scenario, voltage variations of 1.8 V can potentially lead to the polarity inversion of threshold voltage differences (ΔV_{TH0}) in the range of $\pm$ 233.28 µV that is around 1.18% of the distribution showed in Fig. 3.8(b). On the other hand, the temperature scenario is more critical. Indeed, always considering the 3σ scenario, temperature variations of 100 °C can potentially lead to the polarity inversion of threshold voltage differences (ΔV_{TH0}) in the range of $\pm$ 1.53 mV that is around 6.24% of the distribution showed in Fig. 3.8(b). However, referring to the statistical correlations showed in Fig. 3.8(d) and (e) for a given V_{TH0} mismatch the maximum difference in terms of $\lambda_{D1,2}$ and $k_{T1,2}$ is of 70 µV/V and 7.8 µV/K respectively. This means that 1.8 V of voltage variation and 100 °C of temperature variation can potentially lead to the polarity inversion in the V_{TH0} mismatch for ΔV_{TH0} in the range of $\pm$ 126 µV and $\pm$ 780 µV that are the 0.72% and 3.2% of the distribution showed in Fig. 3.8(b) respectively. Moreover, the previous analysis only considers the unstable bits due to the ΔV_{TH0} polarity inversion. Unfortunately, when we consider the PUF solution immersed in a complete system we must also take into account the unstable bits due to on-chip noise (i.e., occasional bit flips due to the on-chip noise). Indeed, noisy bits are function of both core and conversion circuits. For this reason, in the proposed design concept, the conversion block is embedded into the bitcell circuit, thus reducing the impact of noise on the conversion phase.

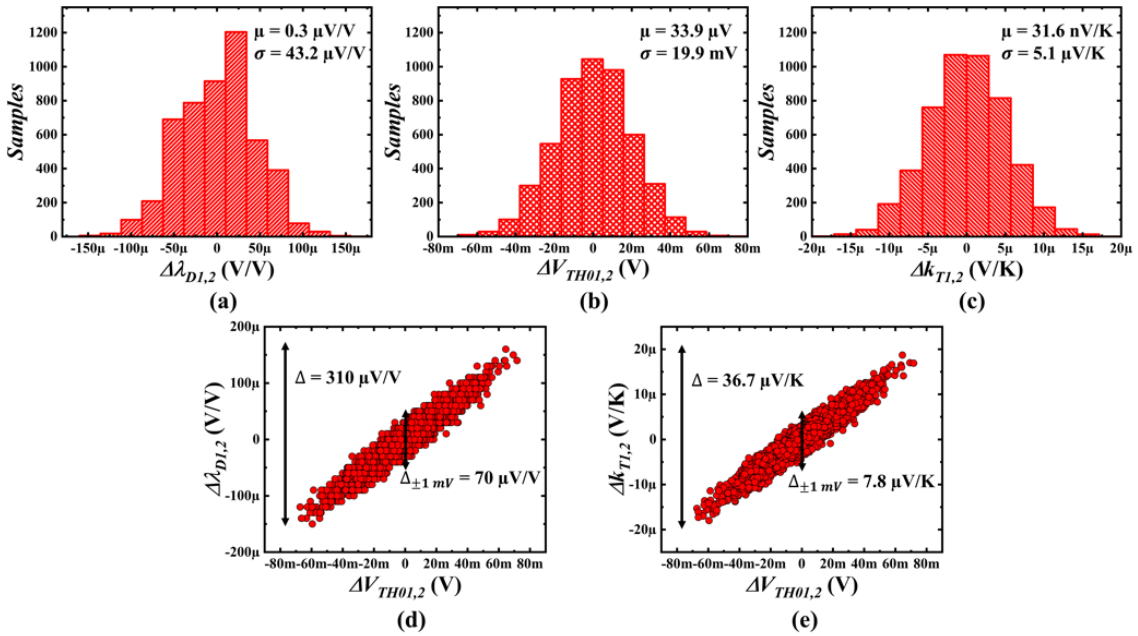
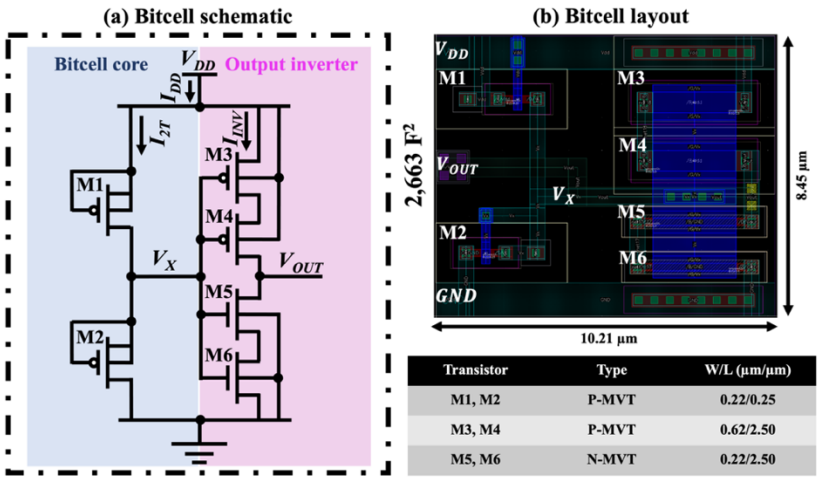


Fig. 3.8. Statistical distributions of (a) the DIBL coefficient difference ($\Delta\lambda_{D1,2}$), (b) the V_{TH0} difference ($\Delta V_{TH1,2}$) and (c) the V_{TH} temperature coefficient difference ($\Delta k_{T1,2}$) for a PMOS MVT device with W = 0.22 μm and L = 0.25 μm (from 5k-run Monte Carlo simulations under local variations in the considered 180-nm technology). Statistical correlation between (d) DIBL coefficient mismatch ($\Delta\lambda_{D1,2}$) and threshold voltage mismatch ($\Delta V_{TH01,2}$) and (e) V_{TH} temperature coefficient mismatch ($\Delta k_{T1,2}$) and threshold voltage mismatch ($\Delta V_{TH01,2}$) for two PMOS MVT devices with W = 0.22 μm and L = 0.25 μm (from 5k-run Monte Carlo simulations under local variations in the considered 180-nm technology).

3.2.4 Simulations results of the 2T-corebased bitcell

Fig. 3.9(a) and (b) illustrates the schematic and layout of the 2T-core based bitcell, respectively.



Transistor	Type	W/L (μm/μm)
M1, M2	P-MVT	0.22/0.25
M3, M4	P-MVT	0.62/2.50
M5, M6	N-MVT	0.22/2.50

Fig. 3.9. (a) Schematic and (b) Layout of the PUF bitcell based on the 2T sub-threshold voltage divider.

In the proposed bitcell solution, showed in Fig. 3.9(a), the conversion block consists of a 4T inverter. Indeed, for reducing the impact of the on-chip noise it is important pushing all the V_X voltages far from the unstable input region of the inverter (i.e., the difference between the

minimum high input voltage, V_{IH}, and the maximum low input voltage, V_{IL}). This is because the not full swing V_X voltage samples that fall in the unstable input region can result potentially unstable at the output of the inverter due to noise. For this reason, we adopted a 4T inverter instead of the conventional 2T in the proposed bitcell, as shown in Fig. 3.9(a). Indeed, the use of four transistors along with a proper sizing ensures obtaining a narrower unstable input region (i.e., lower V_{IH}-V_{IL}) as well as a lower power consumption. This is achieved at the cost of higher bitcell footprint. Indeed, as shown in Fig. 3.9(b), despite the small number of transistors the bitcell area normalized to the adopted technology (i.e., dividing the obtained area by two times the minimum channel length of the technology) is of 2,663F^2 (86.28 µm^2). Fig. 3.10 shows simulation results from 5k-run Monte Carlo simulations at V_{DD}= 1.8 V and T = 25 °C at TT corner. In particular,

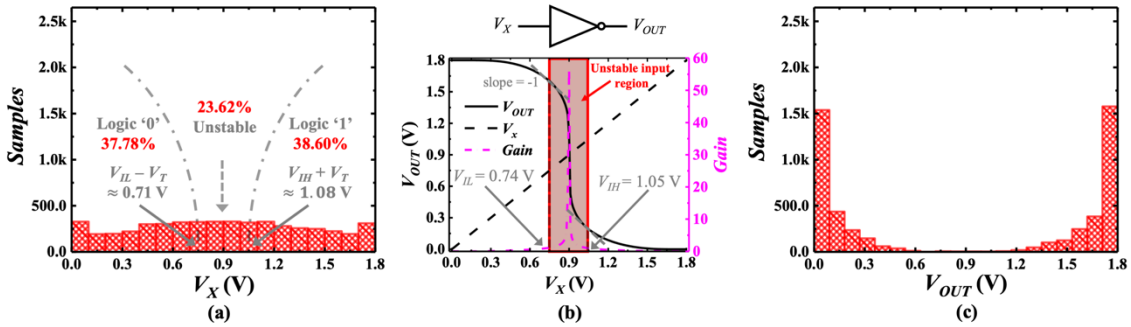


Fig. 3.10. Simulation results (5k-run Monte Carlo at V_{DD} = 1.8 V and T = 25°C) of 2T-core PUF bitcell in 180-nm CMOS at TT corner: (a) statistical distribution of the voltage V_X of the bitcell core, (b) nominal input-output characteristics of the inverter, and (c) statistical distribution of the voltage V_{OUT} of the inverter.

Fig. 3.10(a) illustrates the statistical distribution of the V_X voltage from which the percentage of bits that fall in the unstable region of the inverter plus two thermal voltages (i.e., for taking more effectively into account the effect of noise at different operating conditions), and then potentially unstable, is 23.62% while the percentages of logic '0' and '1' are 37.78% and 38.60%, respectively. Fig. 3.10(b) illustrates the nominal input-output characteristic of the 4T inverter from which the logic threshold (i.e., V_M) is equal to the mid-supply point (i.e., V_{DD}/2) and the unstable region is delimited by minimum high input voltage (V_{IH}) of 1.05 V and a maximum low input voltage (V_{IL}) of 0.74 V. Finally, Fig. 3.10(c) shows the statistical distribution of the voltage V_{OUT} of the inverter. This figure highlights that the trade-off between low DIBL effect and high variability in the 2T-core results in a high percentage of potentially unstable bits at the output of the subsequent inverter. Moreover, Fig. 3.10 also highlights that a single inverter stage is not enough for making the V_X samples full swing thus propagating the noise sensitivity to the readout system. Concerning the performance under PVT variations Fig. 3.11(a)-(d) report the effect of process, voltage, and temperature variations on both amplitude of the unstable input region and input logic threshold of the output inverter. Fig. 3.11(a) and (b) illustrate difference between the minimum high- and maximum low-input voltages (i.e., V_{IH}-V_{IL}) as function of voltage (0.4−1.8 V) and temperature (0−100 °C) variations respectively. From Fig. 3.11(a), such amplitude exhibits a nearly linear decrease with decreasing V_{DD} down to 0.6 V, whereas the decreasing trend deviates from the linearity for V_{DD} below 0.6 V. In addition, Fig. 3.11(b) shows that V_{IH}-V_{IL} slightly increases with increasing the temperature. Moreover, in the corners TT, FS, and SF the static performances of the inverter in terms of V_{IH}-V_{IL} are quite close to each other with an increase of the amplitude at the corner FF and a decrease of the amplitude in the corner SS.

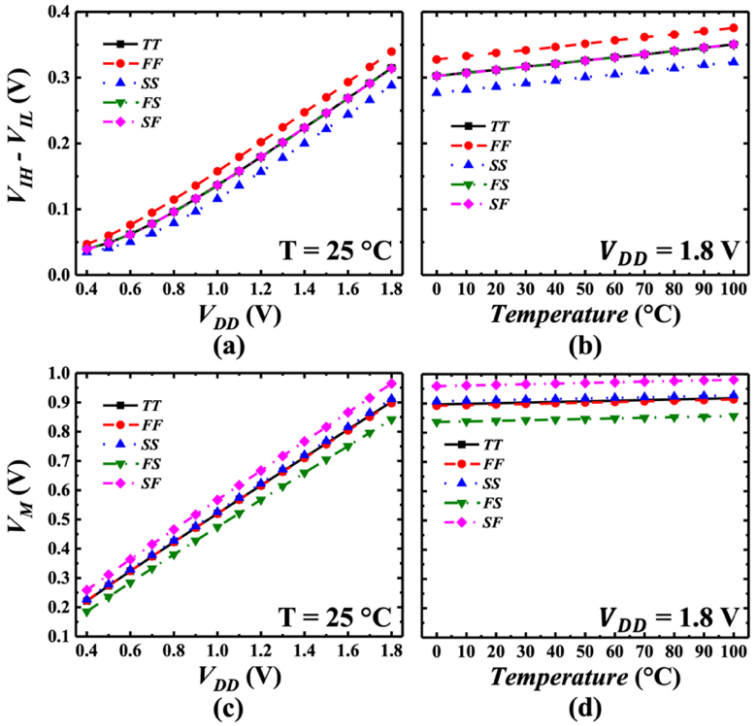


Fig. 3.11. Effect of voltage and temperature variations on the static parameters of the output inverter at different process corners. Effect of (a) voltage with T= 25 °C, and (b) temperature with V_{DD} = 1.8 V variations on the difference between minimum high- and maximum low-input voltages (V_{IH}-V_{IL}). Effect of (c) voltage with T= 25 °C and (d) temperature with V_{DD} = 1.8 V variations on the input logic-threshold (V_M).

From Fig. 3.11(c) the logic threshold shows a linear trend under V_{DD} variations keeping its value quite close to the mid-supply point. From Fig. 3.11(d) the logic threshold exhibits slightly increases when increasing the temperature. Considering the process variations, in both Fig. 3.11(c) and (d) V_M is quite close to the ideal value of $V_{DD}/2$ at the corners TT, FF, SS with a slight increase (decrease) at SF (FS) corner. For better understanding the effect of the process variations it also important understanding their effect on the transistor electrical behavior. Fig. 3.12 report the mean value along with the standard deviation of some electrical parameter such as the threshold voltage ($|V_{TH0}|$), the DIBL coefficient (λ_D), and the threshold voltage temperature coefficient (k_T) for a PMOS MVT device with minimum sizing (i.e., L= 0.25 µm and W= 0.22 µm) at different process corners from 5k-run Monte Carlo simulations.

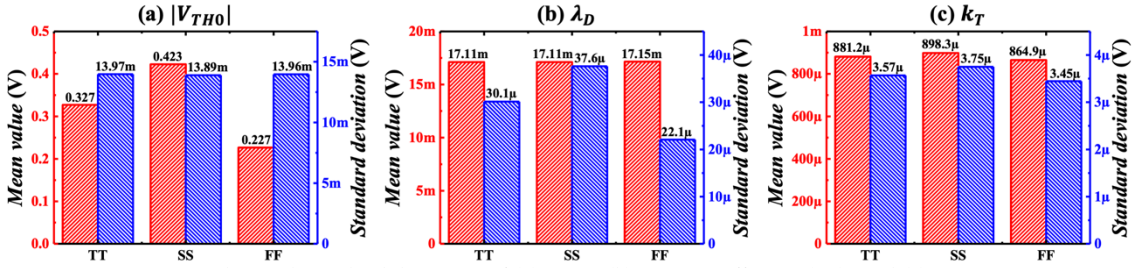


Fig. 3.12. Mean value and standard deviation of (a) V_{TH0}, (b) DIBL coefficient (λ_D), and (c) V_{TH} temperature coefficient (k_T) from 5k-run Monte Carlo simulations at different process corners.

From Fig. 3.12(a) the mean value of $|V_{TH0}|$ at TT corner is of 327 mV which increases up to 422.6 mV at SS corner and decrease down to 226.7 mV at FF corner. On the other hand, the standard deviation slightly variates along the different process corners which could result in a slight lower

V_X variability. From Fig. 3.12(b) the mean value of the DIBL coefficient (λ_D) remains nearly constant (i.e., 17.11 mV/V) under process variations but its standard deviation is 30.11 µV/V at TT corner and increases up to 37.61 µV/V at SS corner and decreases down to 22.08 µV/V at FF corner. This probably leads to higher bitflip under voltage variations at the SS corner. Finally, from Fig. 3.12(c) both the mean value and the standard deviation of the V_{TH} temperature sensitivity (k_T) remain constant under process variations. Fig. 3.13(a)-(c) report the V_X voltage distribution of the 2T-core at TT, FF, and SS corner, respectively, from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T= 25°C). These distributions point out that the adoption of a voltage divider between two nominally identical sub-circuits always guarantee a good uniformity even under process variations. Indeed, the mean value of the distributions is close to the ideal value of the mid-supply point regardless of the considered process corner.

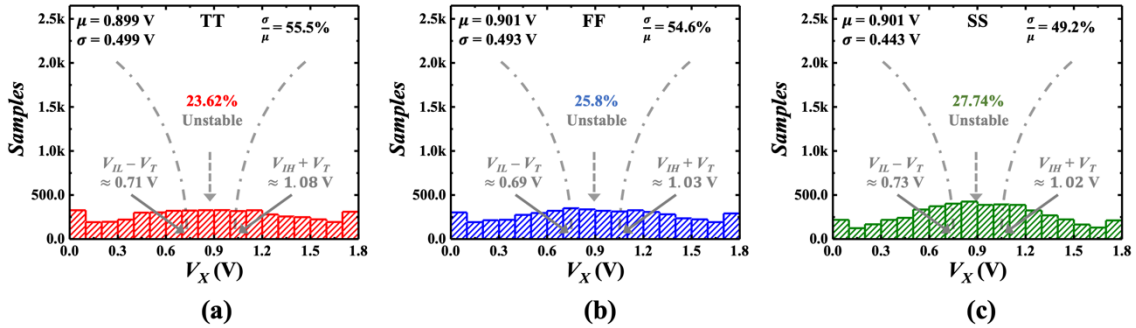


Fig. 3.13. Statistical distribution of the V_X voltage of the 2T-core from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T = 25°C) at (a) TT corner, (b) FF corner, and (c) SS corner.

However, such differences in terms of electrical parameters and their mismatch, showed in Fig. 3.12, lead to different results. Indeed, the distribution at TT corner, Fig. 3.13(a), shows higher V_X spread than that at FF and SS corners thus resulting in a lower percentage of native unstable bits. Concerning the PUF stability at different environmental conditions, Fig. 3.14(a) and (b) provide the percentage of unstable bits under voltage and temperature variations, respectively. In particular, the unstable bits include noisy bits (i.e., bits which flip occasionally due to on chip noise and estimated as the percentage of V_X samples that fall in the unstable input region of the output inverter) and the flipped bits (i.e., V_X samples that cross the input logic threshold, V_M, under voltage and temperature variations).

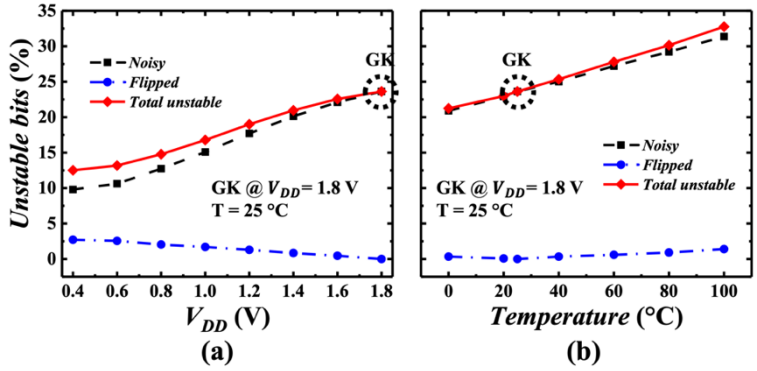


Fig. 3.14. Percentage of simulated unstable bits for the 2T-based solution at TT corner under (a) V_{DD} variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V.

From Fig. 3.14 the overall instability at TT corner is dominated by the noisy bits. In particular, from Fig. 3.14(a) the percentage of bits affected by on-chip noise decreases down to 9.8% when decreasing the V_{DD} down to 0.4 V thanks to the decrease of both the DIBL effect on M1-M2 threshold voltages, which increases the V_X deviation from its nominal value, and the amplitude of the unstable region of the output inverter. Moreover, decreasing the V_{DD} also increases the percentage of flipped bits up to 2.72% due to both DIBL coefficient (λ_D) mismatch and V_M deviation from the mid-supply point. On the other hand, Fig. 3.14(b) show the trend of both noisy and flipped bits under temperature variations. Similar to what happens under voltage variations, the total instability is dominated by noisy bits which increase up to 31.36% when increasing the temperature up to 100 °C while the percentage of flipped bits increases up to 1.4% under the same temperature variation due to both V_{TH} temperature coefficient mismatch (k_T) and V_M deviation from the mid-supply point. However, I would stress again that both DIBL coefficient and V_{TH} temperature coefficient mismatches are underestimated by the circuit simulator. Fig. 3.15(a) and (b) report the percentage of the unstable bits under voltage and temperature variations, respectively, across different process corners.

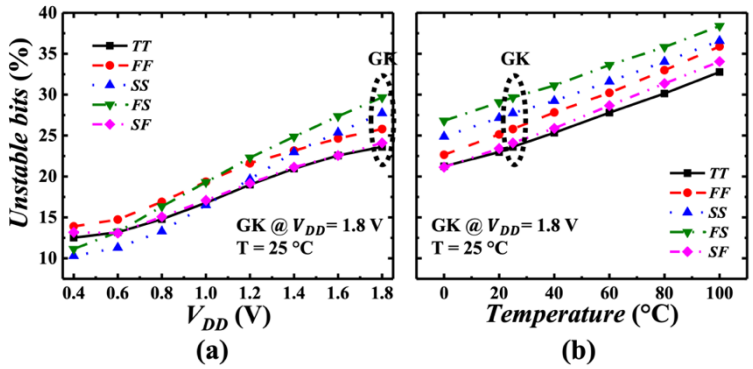


Fig. 3.15. Percentage of total unstable bits for the 2T-based solution across different process corners under (a) voltage variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V.

From this figure the trend is similar to that illustrated in Fig. 3.14. Indeed, the percentage of unstable bits decreases when decreasing the V_{DD} down to 0.4 V and increases when increasing the temperature up to 100 °C. From Fig. 3.15(a) the total instability is lower at TT corner at high V_{DD} while it is lower at SS corner at low V_{DD} this is because: (*i*) the higher DIBL coefficient variability, showed in Fig. 3.12(b) results in higher V_X spread at low V_{DD} and (*ii*) the lower amplitude of the unstable input region of the inverter, showed in Fig. 3.11(a). The different trend at FS and SF corner are strictly related to the output inverter behavior at the same corners. From Fig. 3.15(b) the total instability increases when increasing the temperature up to 100 °C at each corner. Moreover, at TT corner the circuit shows a lower instability compared to that at the other corners. Concerning the power consumption, due to the deep sub-threshold operation the inverter absorbs the most part of the current. In particular, the amount of absorbed current depends on the V_X voltage distribution. Indeed, as closer are to the mid-supply point as higher is the absorbed current by the inverter thus resulting in a higher power consumption. Fig. 3.16(a)-(d) illustrate the absorbed current by both the 2T-core and by the entire bitcell across different process corners and under voltage and temperature variations. Fig. 3.16(a) and (b) report the absorbed current by the 2T-core (I_{2T}) under voltage and temperature variations, respectively.

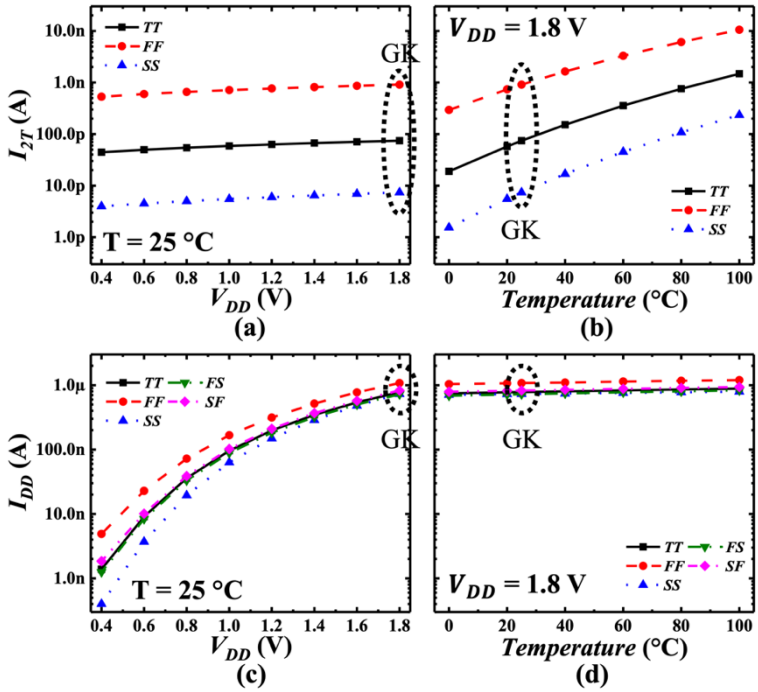


Fig. 3.16. Simulated absorbed current by the 2T-core across different process corners from 5k-run Monte Carlo simulations under (a) voltage variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V. Simulated absorbed currend across different process corners by the bitcell under (c) voltage variations at T= 25°C and (d) temperature variations at V_{DD}= 1.8 V.

From Fig. 3.16(a) the I_{2T} shows a slight deviation from its nominal value under voltage variations. For example, at the TT corner the I_{2T} (P_{2T}) decreases from 74.24 pA (133.63 pW) down to 44.37 pA (17.75 pW) when decreasing the V_{DD} from 1.8 V down to 0.4 V. On the other hand, from Fig. 3.16(b) the same currents exhibit an exponential relationship with the temperature. Indeed, at TT corner when increasing the temperature up to 100 °C the absorbed current (dissipated power) also increases from 74.24 pA (133.63 pW) up to 1.48 nA (2.66 nW). Fig. 3.16(c) and (d) report the supply current of the bitcell (I_{DD}) under voltage and temperature variations respectively. From Fig. 3.16(c) the I_{DD} (P_{DD}) shows an exponential relationship with the V_{DD}. Indeed, at TT corner the I_{DD} (P_{DD}) decrease from 0.78 µA (1.40 µW) down to 1.38 nA (0.55 nW) when decreasing the V_{DD} down to 0.4 V. On the other hand, from Fig. 3.16(d) the absorbed current deviates linearly from its nominal value under temperature variations. Indeed, at TT corner the absorbed current (dissipated power) increases from 0.78 µA (1.40 µW) up to 0.89 µA (1.60 µW) when increasing the temperature up to 100 °C. Considering the process variations, the worst corner is the FF where the higher PMOS conductivity increases the I_{2T} (P_{2T}) and the I_{DD} (P_{DD}) up to 909.28 pA (1.64 nW) and 1.08 µA (1.94 µW), respectively, at GK conditions. On the other hand, the best corner is the SS where the lower PMOS conductivity reduces the I_{2T} (P_{2T}) and the I_{DD} (P_{DD}) down to 7.41 pA (13.34 pW) and 0.73 µA (1.31 µW), respectively, at GK conditions.

3.3 4T Sub-threshold Voltage Divider

The use of a voltage divider between two nominally identical sub-circuits along with a deep sub-threshold operation ensure a high degree of randomness regardless of the supply voltage and

operative temperature. However, along with these benefits the 2T-core based solution shows some limitations:

1. The trade-off between high variability and low DIBL effect limits the V_X voltage spread thus increasing the overall instability as well as the absorbed current by the output inverter.
2. The DIBL coefficient (λ_D) mismatch can potentially lead to a relatively large percentage of flipped bits under voltage variations due to the large voltage drops across M1 and M2.
3. The V_{TH} temperature coefficient (k_T) mismatch can potentially lead to a relatively large percentage of flipped bits under temperature variations.

One possible solution to counteract some of the previous issues consists of adding one negative-V_{SG} transistor (i.e., PMOS-MVT) in each sub-circuit which acts as mismatch boosters.

3.3.1 Operative principle of the 4T voltage divider

Fig. 3.17(a) and (b) illustrate the circuit of the 4T-core bitcell along with the operative principle. The bitcell consists of the 4T-core block along with an output inverter for generating the bit. From Fig. 3.17(a) M1 and M2 act as main mismatch sources determining the mismatch polarity while M3 and M4 act as mismatch boosters pushing the V_X samples toward V_{DD} or ground, depending to the M1-M2 mismatch polarity, as well as shielding M1-M2 against the voltage variations. Indeed, in absence of mismatch the V_X voltage assesses to the mid-supply point (V_{DD}/2) since the two sub-circuits are nominally identical. On the other hand, when mismatch occurs if M1 (M2) is stronger than M2 (M1), the voltage drop across M2 (M1) is higher than that of M1 (M2), thus making M4 (M3) weaker than M3 (M4). This leads to an increase in the voltage drop on the BC (TC), thus pushing V_X toward V_{DD} (ground). This principle is well illustrated in Fig. 3.17(b). The same figure also highlights that the equivalent DIBL of each sub-circuit is much lower than that of the 2T-core, showed in Fig. 3.2(b). As result, for a given mismatch the V_X voltage is pushed further towards V_{DD} or ground compared to the 2T-core solution.

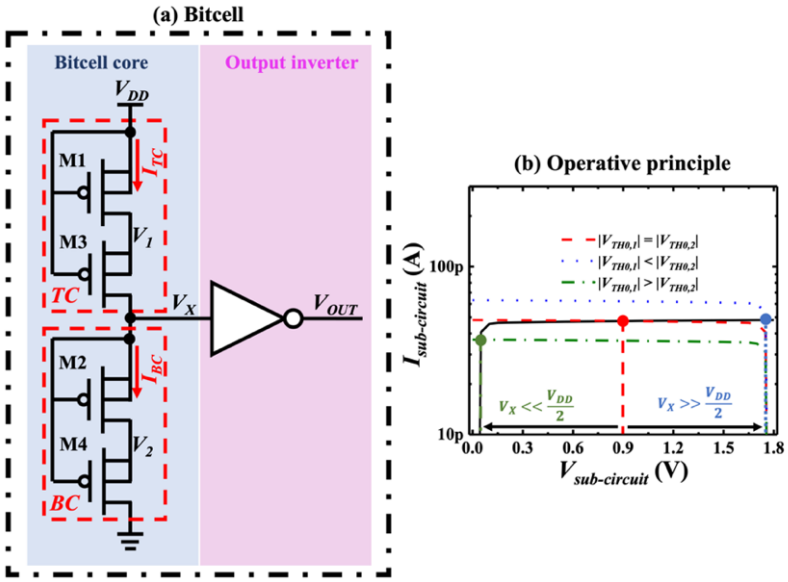


Fig. 3.17. (a) Schematic of the bitcell based on the 4T voltage divider along with (b) the operative principle.

3.3.2 *Design guidelines of the 4T-core*

The main advantage of this circuital approach consists of differently optimizing the two transistors in each sub-circuit. Indeed, M1 and M2 need to be sized for maximizing their variability while M3 and M4 for improving their amplification effect. As result, M1 and M2 translate their strength mismatch into a difference between their voltage drops (i.e., V_{SD1} - V_{SD2}) and then M3 and M4 translate such difference into a difference between their voltage drops (i.e., V_{SD3} - V_{SD4}). This allows increasing the V_X voltage spread as well as keeping nearly constant V_{SD1} - V_{SD2} under voltage variations. However, M3 and M4 need to be properly sized for maximizing this effect. Indeed, they should show the following properties: (*i*) lower variability so that the output bit is function only of M1-M2 mismatch (i.e., the difference between the voltage drops across M1 and M2, V_{SD1} - V_{SD2}, is always higher than the mismatch between M3 and M4, V_{TH4} - V_{TH3}) and (*ii*) enough conductivity (i.e., $W_{3,4}/L_{3,4}$) for ensuring an adequate voltage drop across M1-M2. More precisely, requirement (*ii*) refers to the fact that the voltage drop across M1 and M2 must not be too high for avoiding that the DIBL coefficient (i.e., $\lambda_{D1,2}$) counteracts the M1-M2 mismatch but the at the same time not too low (i.e., at least 3−4 thermal voltages) for maximizing the V_{SD1} - V_{SD2}. The latter point is very critical since for a given mismatch if the operative voltage drop across M1 or M2 (depending on the mismatch) is lower than $3-4V_T$ lower V_{SD1} - V_{SD2} is required for carrying the same current due to the exponential term in the square bracket of (3.1) thus resulting in a lower V_X voltage spread. Fig. 3.18(a)-(c) report the V_X variability (i.e., the ratio between the standard deviation, σ, and the mean value, μ) trend in 180-nm CMOS technology exploiting medium-V_{TH} (i.e., mvt) devices as function of M3-M4 sizing with M1-M2 sized as in the 2T-core solution (i.e., $L_{1,2}$= 0.25 μm and $W_{1,2}$= 0.22 μm) from 1k-run Monte Carlo simulations at TT corner and at GK conditions (i.e., V_{DD}= 1.8 V and T = 25°C). From Fig. 3.18(a) The V_X variability increases up to 78.8% when increasing the M3-M4 channel length ($L_{3,4}$) up to 0.5 μm due to the reduction of the DIBL coefficient. For values above 0.5 μm the V_X variability exhibits a slight decrease due to the reduction of the M1-M2 voltage drops thus also reducing their difference for a given mismatch. On the other hand, from Fig. 3.18(b) the V_X variability strongly increases up to 92.13% when increasing the channel width up to 1.00 μm after which the V_X variability exhibits a slight increase. This can be ascribed to the increase of the voltage drops across M1 and M2 which also improve their difference.

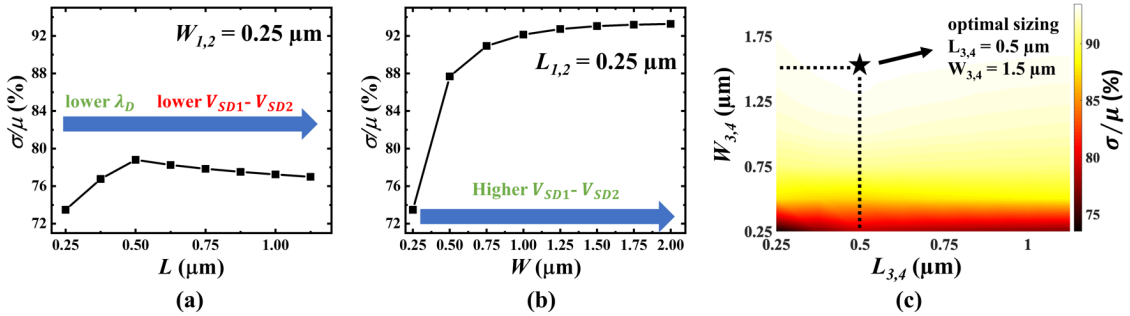


Fig. 3.18. V_X variability as function of (a) $L_{3,4}$ with $W_{3,4}$ = 0.25 μm, (b) $W_{3,4}$ with $L_{3,4}$ = 0.25μm, and (c) both $L_{3,4}$ and $W_{3,4}$ at nominal conditions of V_{DD} = 1.8 V and T = 25 °C from 1k-run Monte Carlo simulations.

This trend is also confirmed by Fig. 3.18(c) where for a given channel length, the increase of the channel width leads to a strong increase of the V_X variability. However, an optimal value also exists for the channel width. Fig. 3.19 reports the trend of the V_X variability as function of $W_{3,4}$.

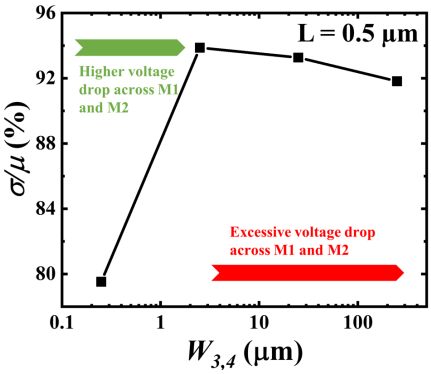


Fig. 3.19. V_X variability as function of the M3-M4 channel widths with L = 0.5 µm at GK conditions (i.e., V_{DD} = 1.8 V and T = 25 °C) from 1k-run Monte Carlo simulations.

From Fig. 3.19 for large $W_{3,4}$ values the M1-M2 voltage drops became too much larger thus leading DIBL effect of M1 and M2 (i.e., $\lambda_{D1,2}$) counteracts their mismatch as in the 2T solution. This results in a decrease of the V_X variability for excessive large channel widths. Indeed, the chosen sizing is reported in Fig. 3.18(c) and consists of $L_{3,4}$= 0.5 µm and $W_{3,4}$= 1.5 µm. Fig. 3.20(a)-(c) illustrate the statistical distributions of the electrical parameters from 5k-run Monte Carlo simulations for a PMOS MVT device with L = 0.5 µm and W = 1.5 µm. Fig. 3.20(a) reports the DIBL coefficient (λ_D) distribution where the mean value is 15.6 mV/V with a standard deviation of 7.4 µV/V thus resulting in overall variability of 0.05%. Fig. 3.20(b) shows the threshold voltage distribution (V_{TH0}), extracted at T = 25 °C and V_{SG}= V_{SD}= V_{SB}= 0, which shows a mean value of -266 mV with a standard deviation of 3.7 mV thus resulting in a variability of 1.39%. Fig.3.20(c) illustrates the threshold voltage temperature coefficient (k_T) from which the mean value is 984.8 µV/K and the standard deviation is 0.45 µV/K, thus resulting in a variability of 0.05%.

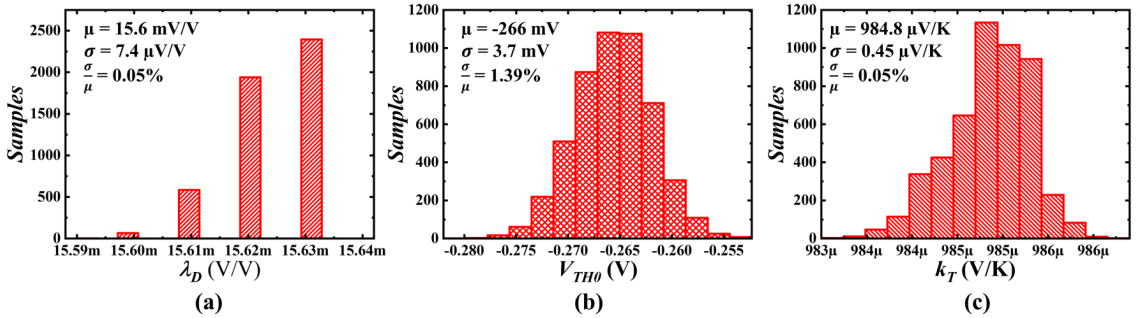


Fig. 3.20. Statistical distribution of (a) DIBL coefficient (λ_D), (b) Threshold voltage (V_{TH0}), and (c) threshold voltage temperature coefficient (k_T) from 5k-run Monte Carlo simulations for a PMOS MVT device with nominal sizing (i.e., L = 0.5 µm and W = 1.5 µm).

From a quantitative point of view, given M1-M4 working in the sub-threshold region their current can be expressed by (3.1). Since M1 is nominally identical to M2 and M3 to M4 we can do the following assumptions: (*i*) n_1= n_2= $n_{1,2}$, (*ii*) $\lambda_{D,1}$= $\lambda_{D,2}$= $\lambda_{D1,2}$, (*iii*) n_3= n_4= $n_{3,4}$, (*iv*) $\lambda_{D,3}$= $\lambda_{D,4}$= $\lambda_{D3,4}$, and (*v*) $k_{T,3}$= $k_{T,4}$= $k_{T3,4}$. Due to the low variability showed in Fig. 3.20 the assumptions

(*iv*) and (*v*) can be considered valid. Moreover, the limited voltage drops across M1 and M2 allows considering valid the assumption (*ii*). Indeed, their difference is close enough for making infrequent the possibility of flipping under voltage variations. From (3.1) and Fig. 3.17, the above assumption and considering the voltage drop across each transistor higher than 3-4V_T, the currents of M1-M4 are given by

$$I_{M1} = I_{0,1}\frac{W_1}{L_1}\exp\left(\frac{V_{TH0,1}(T) + \lambda_{D1,2}(V_{DD} - V_1)}{n_{1,2}V_T}\right) \quad (3.7)$$

$$I_{M2} = I_{0,2}\frac{W_2}{L_2}\exp\left(\frac{V_{TH0,2}(T) + \lambda_{D1,2}(V_X - V_2)}{n_{1,2}V_T}\right) \quad (3.8)$$

$$I_{M3} = I_{0,3}\frac{W_3}{L_3}\exp\left(\frac{V_1 - V_{DD} + V_{TH0,3}(T) + \lambda_{D3,4}(V_1 - V_X)}{n_{3,4}V_T}\right) \quad (3.9)$$

$$I_{M4} = I_{0,4}\frac{W_4}{L_4}\exp\left(\frac{V_2 - V_X + V_{TH0,4}(T) + \lambda_{D3,4}(V_2)}{n_{3,4}V_T}\right) \quad (3.10)$$

By equating (3.7) and (3.9) (i.e., the currents of M1 and M3), and (3.8) and (3.10) (i.e., the currents of M2 and M4), the voltages V_1 and V_2 are given by

$$V_1 = \frac{1}{n_{1,2} + n_{1,2}\lambda_{D3,4} + n_{3,4}\lambda_{D1,2}}\left[\left(n_{1,2} + n_{3,4}\lambda_{D1,2}\right)V_{DD} + n_{3,4}V_{TH0,1}(T) - n_{1,2}V_{TH0,3}(T) + n_{1,2}\lambda_{D3,4}V_X + n_{1,2}n_{3,4}V_T\ln\left(\frac{I_{0,1}W_1L_3}{I_{0,3}W_3L_1}\right)\right] \quad (3.11)$$

$$V_2 = \frac{1}{n_{1,2} + n_{1,2}\lambda_{D3,4} + n_{3,4}\lambda_{D1,2}}\left[n_{3,4}V_{TH0,2}(T) - n_{1,2}V_{TH0,4}(T) + \left(n_{1,2} + n_{3,4}\lambda_{D1,2}\right)V_X + n_{1,2}n_{3,4}V_T\ln\left(\frac{I_{0,2}W_2L_4}{I_{0,4}W_4L_2}\right)\right] \quad (3.12)$$

Then, by substituting (3.11) and (3.12), respectively, in (3.7) and (3.8) and equating the resulting expressions, the equation for the voltage V_X can be derived as follow.

$$V_X = \frac{V_{DD}}{2} + \frac{1}{2\lambda_{D3,4}}\left[V_{TH0,3}(T) - V_{TH0,4}(T) + n_{3,4}V_T ln\left(\frac{I_{0,3}W_3L_4}{I_{0,4}W_4L_3}\right)\right] + \frac{1 + \lambda_{D3,4}}{2\lambda_{D1,2}\lambda_{D3,4}}\left[V_{TH0,1}(T) - V_{TH0,2}(T) + n_{1,2}V_T ln\left(\frac{I_{0,1}W_1L_2}{I_{0,2}W_2L_1}\right)\right] \quad (3.13)$$

If we also explicit the V_{TH} temperature sensitivity of M1 and M2 the equation (3.13) can be written as follow

$$V_X = \frac{V_{DD}}{2} + \frac{1}{2\lambda_{D3,4}}\left[V_{TH0,3} - V_{TH0,4} + n_{3,4}V_T ln\left(\frac{I_{0,3}W_3L_4}{I_{0,4}W_4L_3}\right)\right] + \frac{1 + \lambda_{D3,4}}{2\lambda_{D1,2}\lambda_{D3,4}}\left[V_{TH0,1} - V_{TH0,2} + \left(k_{T,1} - k_{T,2}\right)(T - T_{nom}) + n_{1,2}V_T ln\left(\frac{I_{0,1}W_1L_2}{I_{0,2}W_2L_1}\right)\right] \quad (3.14)$$

Where the low M3-M4 variability, showed in Fig. 3.20(c), allows us neglecting the temperature sensitivity of these transistors.

Moreover, the low V_{TH0} variability showed in Fig. 3.20(b) allows us approximating the equation (3.14) as follow

$$V_X \approx \frac{V_{DD}}{2} + \frac{1+\lambda_{D3,4}}{2\lambda_{D1,2}\lambda_{D3,4}}\left[V_{TH0,1} - V_{TH0,2} + \left(k_{T,1} - k_{T,2}\right)(T - T_{nom}) + n_{1,2}V_T ln\left(\frac{I_{0,1}W_1L_2}{I_{0,2}W_2L_1}\right)\right] \quad (3.15)$$

From equation (3.15) in absence of mismatch $V_{TH0,1} = V_{TH0,2}$ and $V_X = V_{DD}/2$. However, due to the mismatch, differences in $V_{TH0,1} - V_{TH0,2}$ are translated into V_X deviation from the mid-supply point as far as low is the equivalent DIBL of the sub-circuit (i.e., $\lambda_{D1,2}\lambda_{D3,4}$). Similar to the 2T solution, the second part of equation (3.15) refers to the temperature sensitivity. Indeed, mismatch in the k_T terms can lead to a bit flip (i.e., inversion of the mismatch polarity, $V_{TH0,1} - V_{TH0,2}$) at a certain temperature. Assuming V_{TH0} variations of M1 and M2 as statistically independent in (3.15) and neglecting the logarithm term, the standard deviation of V_X (i.e., σ_{V_X}) is given by

$$\sigma_{V_X} \approx \frac{\sigma_{V_{TH01,2}}}{\sqrt{2}\lambda_{D1,2}\lambda_{D3,4}} \quad (3.16)$$

where $\sigma_{V_{TH01,2}}$ is the V_{TH0} standard deviation of M1 and M2. Typically, λ_D values are $\ll 1$, especially for long-channel devices, thus resulting in a higher gain and then a higher V_X spread compared to (3.6). To better understand the improvement with respect to the 2T core solution Fig. 3.21(a) and (b) illustrate the V_X voltages as function of the M1-M2 mismatch in terms of $V_{TH0,1} - V_{TH0,2}$ and the 4T-core gain distribution over the M1-M2 mismatch, respectively, from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T = 25°C).

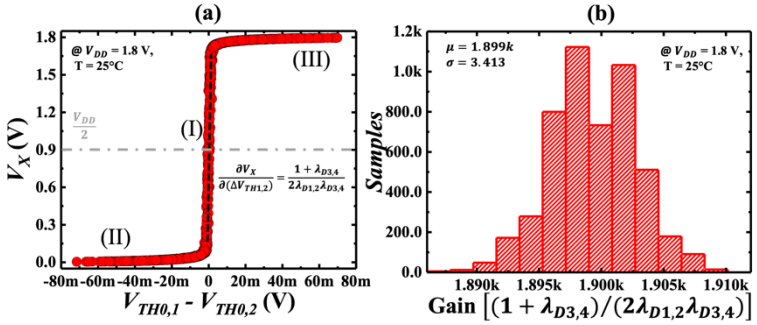


Fig. 3.21. (a)V_X Voltage as function of the M1-M2 mismatch at GK conditions (i.e., V_{DD} = 1.8 V and T = 25°C) and (b) gain (i.e., $(1+\lambda_{D3,4})/(2\lambda_{D1,2}\lambda_{D3,4})$) distribution from 5k-run Monte Carlo simulations.

Also here, the analytical treatment developed above well describes the region (I) of Fig.3.21(a) in which the assumption made above (i.e., the voltage drops across the two transistors are at least 3−4 thermal voltages for making the terms in square bracket of (3.1) negligible) is valid. Indeed, the DIBL coefficients of M1-M2 (i.e., $\lambda_{D1,2}$) and M3-M4 (i.e., $\lambda_{D3,4}$) are equal to 17.1 mV/V and 15.6 mV/V respectively thus resulting in an average overall gain (i.e., $1/2\lambda_{D1,2}\lambda_{D3,4}$) around 1900, as shown in Fig. 3.21(b). In region (II) and (III) the M1-M2 mismatch is large enough for letting the strongest transistor among them works with a V_{SD} lower than 3−4 thermal voltages thus approximately saturating the V_X voltages towards V_{DD} or ground according to the M1-M2 mismatch. Like the 2T-core solution the slope of the characteristic in (II) and (III) is not zero but it is much lower than that in region (I). This is because the exponential term in the square bracket leads to large current variations with small voltage (i.e., $V_{SD1,2}$) variations. In this way, a small

voltage variation is required on the strongest transistor for compensating the strength difference caused by the mismatch.

For better understanding the M3-M4 sizing, equating (3.7) and (3.8) and neglecting the resulting logarithmic term the difference between the voltage drops across M1 and M2 can be expressed as follow.

$$V_{SD1} - V_{SD2} \approx \frac{V_{TH0,2}(T) - V_{TH0,1}(T)}{\lambda_{D1,2}} \quad (3.17)$$

On the other hand, equating (3.9) and (3.10) the difference between the voltage drops across M3 and M4 is given by

$$V_{SD3} - V_{SD,4} \approx \frac{V_{TH0,4}(T) - V_{TH0,3}(T)}{\lambda_{D3,4}} + \frac{V_{SD1}(T) - V_{SD2}(T)}{\lambda_{D3,4}} \approx \frac{V_{TH0,4}(T) - V_{TH0,3}(T)}{\lambda_{D3,4}} + \frac{V_{TH0,2}(T) - V_{TH0,1}(T)}{\lambda_{D1,2}\lambda_{D3,4}} \quad (3.18)$$

As discussed above, M3-M4 sizing must reduce their variability as well as ensure adequate voltage drops across M1 and M2 so that for a given mismatch the second term in (3.18), $(V_{SD1} - V_{SD2})/\lambda_{D3,4}$, is always higher than the first term, $(V_{TH0,4} - V_{TH0,3})/\lambda_{D3,4}$. We need also consider that large voltage drops across M3 and M4 could emphasize the opposite mismatch (i.e., $V_{TH0,4} - V_{TH0,3}$) due to their DIBL effect. However, this effect rarely affects the output and can be counteracted by using long channel devices (i.e., with low DIBL effect coefficient). As result, the V_X voltage is function only of the M1-M2 mismatch. Concerning the latter point Fig. 3.22 shows the percentage of samples in which the output bit polarity is function of M1-M2 mismatch (i.e., samples in which the absolute value of $V_{SD1} - V_{SD2}$ is lower than that of $V_{TH0,4} - V_{TH0,3}$) as function of $W_{3,4}$.

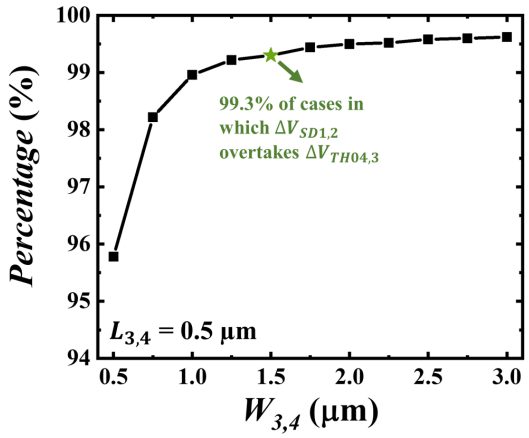


Fig. 3.22. Percentage of samples in which M1-M2 mismatch overtakes the M3-M4 mismatch as function of $W_{3,4}$ sizing with $L_{3,4}$ = 0.5 µm at GK conditions (i.e., V_{DD} = 1.8 V and T = 25 °C) from 5k-run Monte Carlo simulations.

Fig. 3.22 highlights how increasing the M3-M4 conductivity emphasizes the M1-M2 mismatch instead of that of M3-M4 thus improving the probability of generating an output bit function only of M1-M2 mismatch. Indeed, for the chosen $W_{3,4}$ value (i.e., 1.5 µm) this probability is equal to 99.28%. For better understanding the circuit behavior, Fig. 3.23 shows the voltage repartition

normalized to the V_{DD} in the voltage divider for both cases of large, (a) and (c), and small mismatch, (b) and (d).

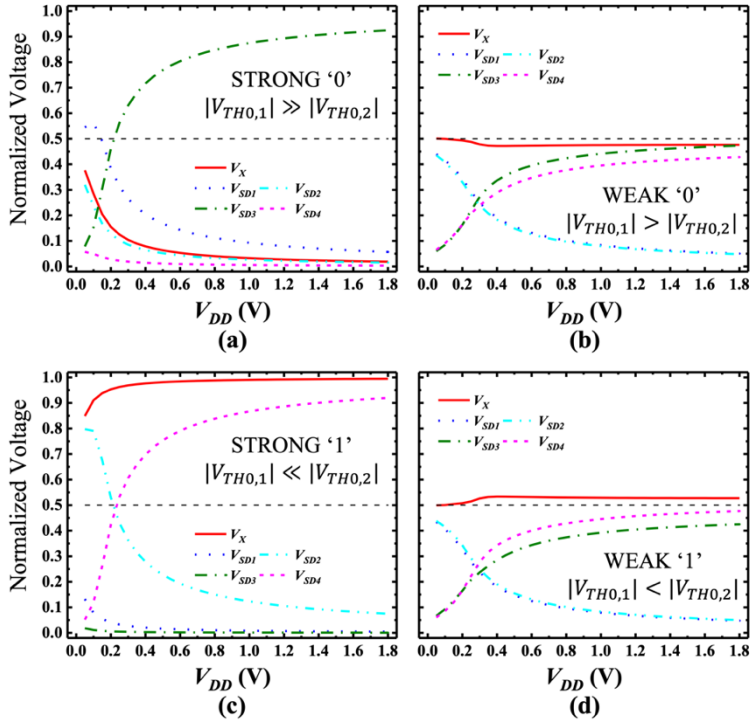


Fig. 3.23. Voltage drops across M1-M4 transistors in the 4T voltage divider normalized to the V_{DD} for strong M1-M2 mismatch and weak M1-M2 mismatch. (a) Strong logic '0', (b) weak logic '0', (c) strong logic '1', and (d) weak logic '1'.

Fig. 3.23(a) and (c) illustrate the voltage drops across each transistor in the voltage divider in the case of strong '0' and '1', respectively. These figures highlight the nearly constant $V_{SD,1} - V_{SD,2}$ trend under voltage variations. Moreover, it is also notable the high gain provided by M3 and M4 even with small mismatch (i.e., with $V_{SD,1} - V_{SD,2}$). Indeed, Fig. 3.23(b) and (d) report the voltage drops across these transistors in the case of weak '0' and '1', respectively. These figures highlight the ability of the 4T-core of pushing the V_X voltage far from the mid-supply point even when the M1-M2 mismatch is small. These figures also highlight the shielding effect provided by M3 and M4 which keep the voltage drops across M1 ($V_{SD,1}$) and M2 ($V_{SD,2}$) constant under voltage variations thus avoiding potential flips due to the mismatch of their DIBL coefficients.

3.3.3 Simulations and measurements of the 4T-core

To better understand the circuit behavior under VT variations Fig. 3.24 and Fig. 3.25 illustrates the simulated and measured V_X and I_{4T} (i.e., the absorbed current by the core circuit) trends under voltage and temperature variations, respectively. In particular, measurements of the 4T bitcell core were performed following the same procedure as for the 2T core solution (i.e., at wafer level across 20 samples. Fig. 3.24(a) and (b) show the trend of the V_X normalized to V_{DD} simulated and measured across voltages at T = 25 °C from 250 samples and 20 dice respectively.

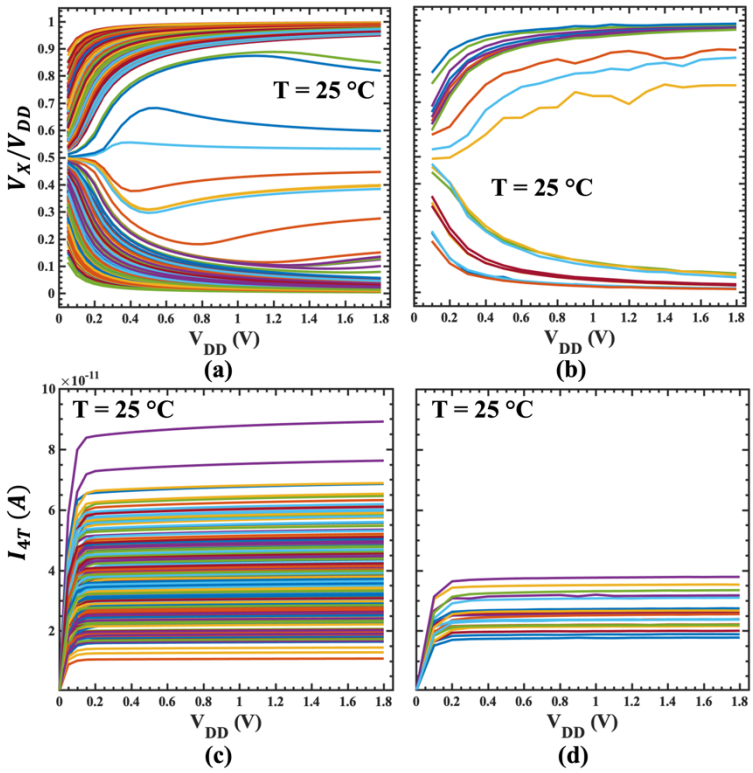


Fig. 3.24. V_X voltage of the 4T-core normalized to V_{DD} (a) simulated across voltages at T = 25°C from 250 samples and (b) measured across voltages at T = 25°C from 20 samples. Absorbed current from the 4T core (i.e., I_{4T}) (c) simulated across voltages at T = 25°C from 250 samples and (d) measured across voltages at T = 25°C from 20 dice.

Fig. 3.24(a) highlights that the amplification effect provided by the circuit is enough large for pushing all the V_X voltages far from the mid-supply point except for very infrequent cases in which the M1-M2 mismatch is very small. This is confirmed by the measurements in Fig. 3.24(b) which also highlights that the shielding effect provided by M3 and M4 counteracts the effect of the DIBL coefficient mismatch of M1 and M2 thus dramatically improving the resilience to the voltage variations. Fig. 3.24(c) and (d) show the trend of the absorbed current (i.e., I_{4T}) simulated and measured across voltages at T = 25 °C from 250 samples and 20 dice, respectively. From Fig. 3.24(c) it is notable how above a certain V_{DD} value the absorbed current exhibits a nearly constant trend under voltage variations which is confirmed by the measurements as shown in Fig. 3.24(d) thus proving the very low sensitivity to the voltage variations. Indeed, from Fig. 3.24(d) the 4T-core supply current (dissipated power) averaged over 20 samples decreases from 25.83 pA (46.49 pW) down to 25.29 pA (10.12 pW) when decreasing the V_{DD} from 1.8 V down to 0.4 V. On the other hand, Fig. 3.25(a) and (b) show the trend of the V_X normalized to V_{DD} simulated and measured across temperatures at V_{DD}= 1.8 V from 250 samples and 20 dice, respectively. From Fig. 3.25(a) the V_X voltages show a nearly temperature-independent trend for samples close to the edges and a linear relationship for sample close to the mid-supply point. Like the 2T-core solution the V_X sensitivity to the temperature variations is strictly related to the M1-M2 mismatch in terms of the V_{TH} temperature coefficient (k_T). Indeed, for the samples that are close to the mid-supply point the linear relationship can be modeled following equation (3.15) for which the temperature variations are amplified by a factor equal to $[(k_{T,1} - k_{T,2})(1 + \lambda_{D3,4})]/(2\lambda_{D1,2}\lambda_{D3,4})$. However, the probability of having a bit flip is only associated to the $k_{T,1} - k_{T,2}$ (i.e., like the 2T-core solution). The reason for which the V_X samples close to the edges show a nearly constant trend

under temperature variation is that until the temperature variations are large enough for reducing enough the M1-M2 mismatch (i.e., $V_{TH0,1} - V_{TH0,2}$) so much that the circuit operates in the region (I) of Fig. 3.21, the V_{TH0} difference of M1 and M2 is large enough for being pushed far from the mid-supply point by the 4T-core.

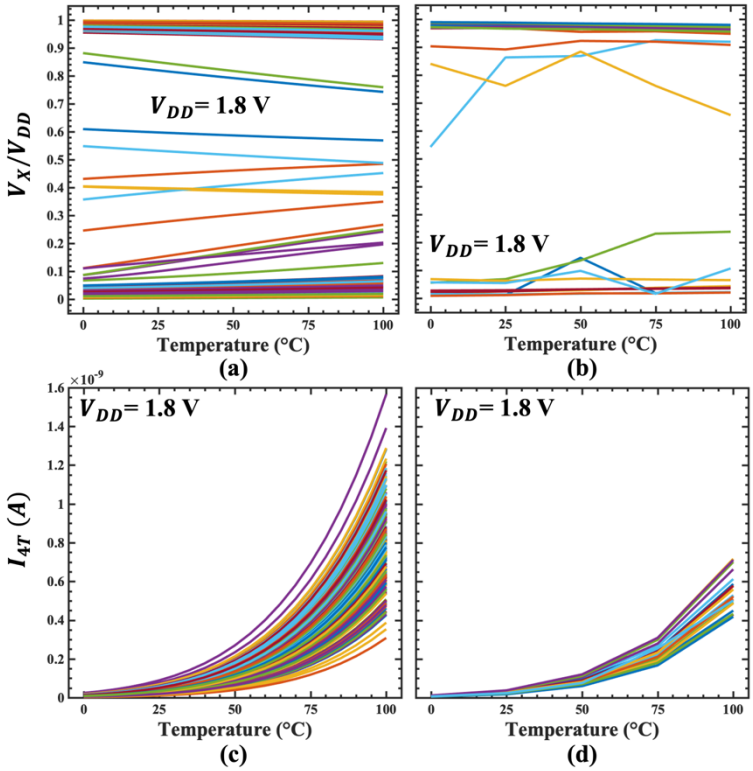


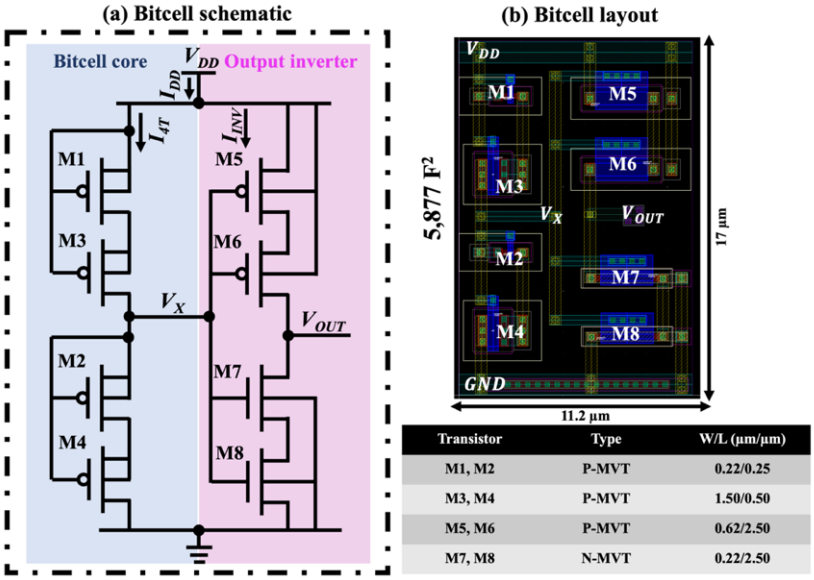
Fig.3.25. V_X voltage of the 4T-core normalized to V_{DD} (a) simulated across temperatures at V_{DD}= 1.8 V from 250 samples and (b) measured across temperatures at V_{DD}= 1.8 V from 20 samples. Absorbed current from the 4T core (i.e., I_4T) (c) simulated across temperatures at V_{DD}= 1.8 V from 250 samples and (d) measured across temperatures at V_{DD}= 1.8 V from 20 dice.

Fig. 3.25(c) shows an exponential relationship between the absorbed current and the temperature. This trend is also confirmed by measurements in Fig. 3.25(d), and it is due to the deep sub-threshold operation. Indeed, from Fig. 3.25(d) the 4T-core supply current (dissipated power) averaged over 20 samples increases from 25.83 pA (46.49 pW) up to 550.26 pA (990.47 pW) when increasing the temperature from 25°C up to 100°C.

3.3.4 Simulation results of the 4T-core based bitcell

Fig. 3.26(a) and (b) illustrate the schematic and layout, respectively, of the PUF bitcell based on the 4T sub-threshold voltage divider along with the adopted sizing. The 4T based bitcell, showed in Fig. 3.26(a), consists of the 4T-voltage divider described in this section along with the same output inverter of Fig. 3.9. Using the same high gain inverter ensure a narrower unstable input region and then a lower probability of having an unstable output bit. From Fig. 3.26(b) the area occupied by this solution is of $5{,}877F^2$ which is more than the double of that occupied by the 2T-core based bitcell. This worsening of area is strictly related to the connection of the transistor

body terminals of the 4T-core block to the respective source terminals thus implying a minimum distance between two consecutive n-well.



Transistor	Type	W/L (µm/µm)
M1, M2	P-MVT	0.22/0.25
M3, M4	P-MVT	1.50/0.50
M5, M6	P-MVT	0.62/2.50
M7, M8	N-MVT	0.22/2.50

Fig. 3.26. (a) Schematic and (b) Layout of the PUF bitcell based on the 4T sub-threshold voltage divider.

Fig. 3.27 shows simulation results at TT corner of both 2T-core and 4T-core based bitcells from 5k-run Monte Carlo simulations at V_{DD}= 1.8 V and T = 25 °C.

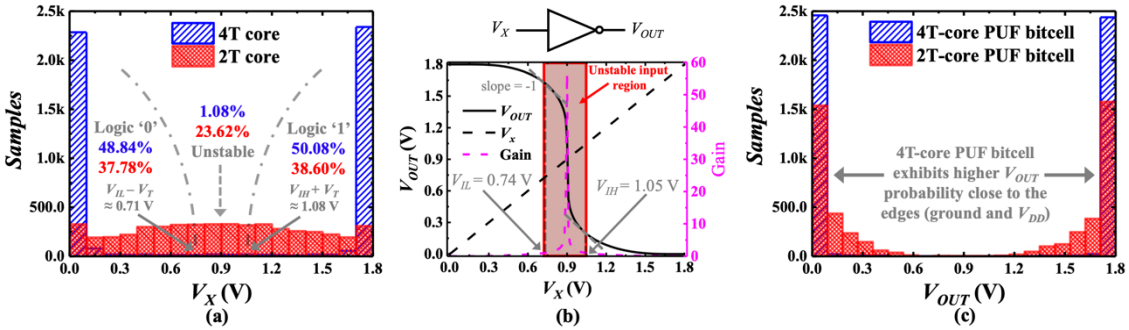


Fig. 3.27. Simulation results (5k-run Monte Carlo at V_{DD} = 1.8 V and 25 °C) of 2T-core versus 4T-core PUF bitcell in 180-nm CMOS: (a) statistical distribution of the voltage V_X of the bitcell core, (b) nominal input–output characteristics of the inverter, and (c) statistical distribution of the voltage V_{OUT} of the inverter.

From Fig. 3.27(a) V_X voltage of the 4T-core shows a nearly full swing deviation due to the amplification effect provided by M3 and M4. This results in a decrease of unstable bits (i.e., bits that fall in the unstable input region plus two thermal voltages of the subsequent inverter) down to 1.08%, which correspond to statistically infrequent case of extremally small M1-M2 mismatch. As result of higher V_X dispersion achieved by the 4T-core, the corresponding distribution of the output voltage V_{OUT} of the inverter shown in Fig. 3.27(c) exhibits a higher probability close to the edges of the voltage range (i.e., ground and V_{DD}). Before showing the circuit behavior under PVT variations it is important to understand what happens to the transistors in the voltage divider at different process corners. Fig. 3.28(a)-(c) report the mean value of threshold voltage ($|V_{TH0}|$), the DIBL coefficient (λ_D), and the threshold voltage temperature coefficient (k_T) for both M1-M2 (i.e., L= 0.25 µm and W= 0.22 µm) and M3-M4 (i.e., L= 0.5 µm and W= 1.5 µm) at different process corners from 5k-run Monte Carlo simulations.

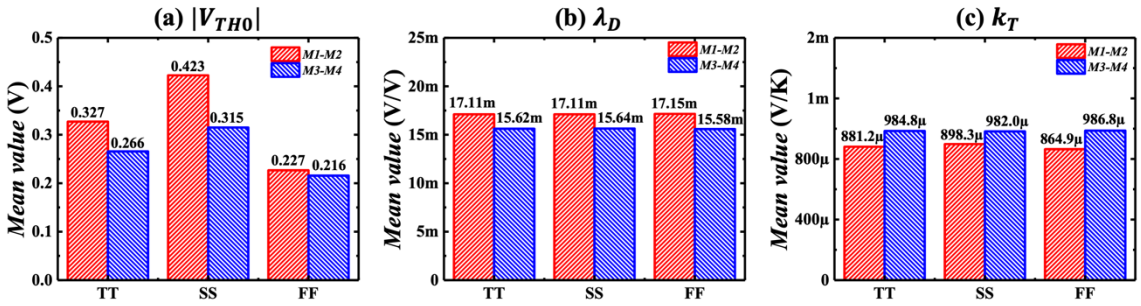


Fig.3.28. Mean value of (a) V_{TH0}, (b) DIBL coefficient (λ_D), and (c) V_{TH} temperature coefficient (k_T) of M1-M2 and M3-M4 transistors from 5k-run Monte Carlo simulations at different process corners.

From Fig. 3.28(a) the mean value of the threshold voltages for M1-M2 and M3-M4 devices at TT are 0.327 V and 0.266 V, respectively. At SS corner the M1-M2 and M3-M4 threshold voltages increase of 29.36% (e.g., 0.423 V) and 18.42% (e.g., 0.315), respectively, of their values at TT corner. On the other hand, at FF corner the M1-M2 and M3-M4 threshold voltages decrease of 30.58% (e.g., 0.227 V) and 18.8% (e.g., 0.216 V) of their values at TT corner. This can be ascribed to the fact that, referring to the core block, at these corners the variation in PMOS strength is more pronounced in minimum-sized M1-M2 than M3-M4. From Fig. 3.28(b) the DIBL coefficient exhibits a nearly constant trend under process variations with mean values of 17.11 mV/V and 15.62 mV/V for M1-M2 and M3-M4, respectively. Finally, from Fig. 3.28(c) the V_{TH} temperature coefficients of M1-M2 and M3-M4 are 881.2 µV/K and 984.8 µV/K, respectively, at TT corner. At SS corner, the k_T coefficient increases of 1.94% (e.g., 898.3 µV/K) for M1-M2 and decreases of 0.28% (e.g., 982.0 µV/K) for M3-M4. At FF corner, k_T decreases of 1.85% (e.g., 864.9 µV/K) for M1-M2 and increases of 0.20% (e.g., 986.8 µV/K) for M3-M4. Fig. 3.29(a)-(c) report the V_X voltage distribution of the 4T-core from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T= 25°C) at TT, FF, and SS corner, respectively.

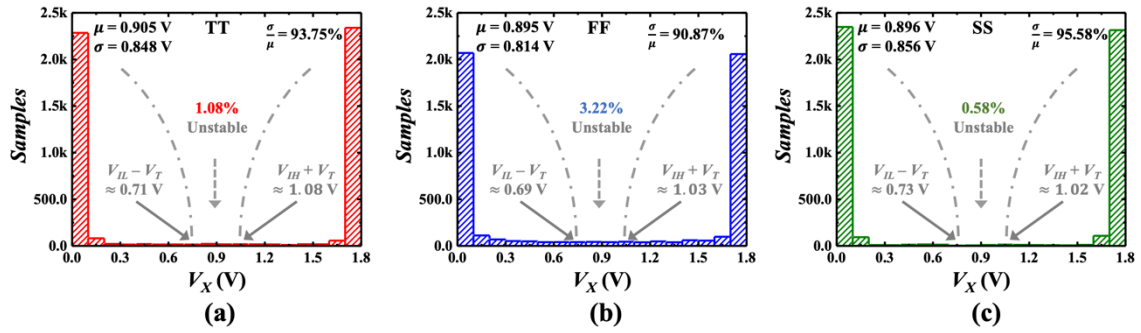


Fig. 3.29. Statistical distribution of the V_X voltage of the 4T-core from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T = 25°C) at (a) TT corner, (b) FF corner, and (c) SS corner.

Like to the 2T-core solution the adoption of a voltage divider between two nominally identical sub-circuits ensures a good uniformity regardless of the process variations as showed in Fig. 3.29. Indeed, the mean values of the distributions at TT, FF, and SS corners are quite similar and very close to the ideal value of 0.9 V. However, variations in the transistor electrical properties lead to variations in the circuit performance across different process corners. For example, from the V_X distribution at TT in Fig. 3.29(a) the standard deviation is 0.848 V thus resulting in a variability of 93.75%. This variability increases up to 95.58% at SS corner while decreases down to 90.87%

at FF corner. This variability variation affects the overall stability. Indeed, at TT corner the percentage of unstable bits is 1.08% which decreases down to 0.58% at SS corner and increase up to 3.22% at FF corner. This can be ascribed to the fact that smaller transistors (i.e., M1 and M2) are more sensitive to the process variations than the larger transistors (i.e., M3 and M4) as shown in Fig. 3.28. In this way, at SS corner the lower M1-M2 conductivity increases their voltage drops thus resulting in higher $V_{SD1} - V_{SD2}$ and hence in a higher V_X spread. On the other hand, at FF corner the higher M1-M2 conductivity decreases their voltage drops thus reducing their difference and hence the V_X spread. As regard, Fig. 3.30 provides the mean value of $|V_{SD1} - V_{SD2}|$ and $|V_X - V_{DD}/2|$ of the 4T-core circuit at different process corners from 5k-run Monte Carlo simulations at GK conditions.

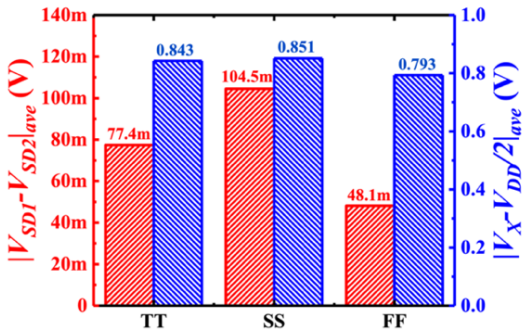


Fig.3.30. Mean values of $|V_{SD1} - V_{SD2}|$ and $|V_X - V_{DD}/2|$ of the 4T-core at different process corners from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C).

Fig. 3.30 highlights that at SS corner the average difference between the voltage drops across M1 and M2 is higher than that at TT corner thus resulting in an increase of the V_X deviation from the mid-supply point and hence in a lower percentage of unstable bits. On the other hand, at FF corner the higher M1-M2 conductivity is translated into a lower average difference between their voltage drops thus resulting in a lower V_X voltage deviation from its nominal value and hence in a higher percentage of unstable bits. Concerning the effect of VT variations on the overall stability, Fig. 3.31(a) and (b) report the simulated unstable bits as function of voltage (with T= 25 °C) and temperature (with V_{DD}= 1.8 V) variations at TT corner, respectively, from 5k-run Monte Carlo simulations. From this figure, the overall instability was reduced of one decade across all the considered voltages and temperature compared to the 2T-core solution. In particular, from Fig. 3.31(a) the instability decreases when decreasing the V_{DD} mainly due to the reduction of the unstable region of the inverter as shown in Fig. 3.11(a). However, the overall instability under voltage variations is dominated by noisy bits. Indeed, when decreasing the V_{DD} the percentage of unstable bits decreases down to 0.8% of which only the 0.14% is composed by flipped bits thus indicating the effectiveness of the shielding effect provided by M3 and M4. Moreover, the cases in which bitflips occur under voltage variations are associated to very small M1-M2 mismatch and can be ascribed to the DIBL coefficient mismatch between M1 and M2 and to the fact that, when decreasing the V_{DD} their strength difference in terms of $V_{SD,1} - V_{SD,2}$ also decreases and this became crucial if they operate with voltage drops lower than $3-4V_T$. Indeed, considering equation (3.18), at very low voltages M1 and M2 may operate with $V_{SD} < 3-4V_T$ and their strength difference in terms of $V_{SD,1} - V_{SD,2}$ may be lower than M4-M3 mismatch (i.e., $V_{TH0,4} -$

$V_{TH0,3}$) so that the latter may determinate the output bits. However, the increase of the V_{DD} may push M1 and M2 to operate with $V_{SD} > 3{-}4V_T$ so that their strength difference became higher than M4-M3 mismatch thus determining the output bit and changing its polarity with respect to the value observed at very low voltages. Anyway, this happens in infrequent cases.

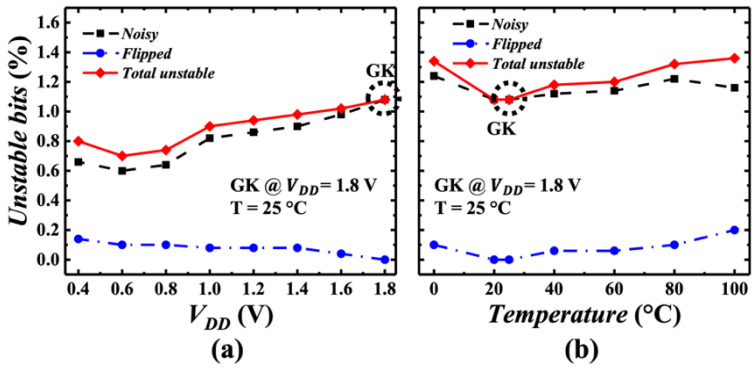


Fig. 3.31. Percentage of simulated unstable bits for the 4T-based solution from 5k-run Monte Carlo simulations at TT corner under (a) V_{DD} variations at T= 25 °C and (b) temperature variations at V_{DD} = 1.8 V.

On the other hand, the instability increases when varying the temperature. In particular, when increasing the temperature up to 100 °C due to the reduction of both the M1-M2 strength and the amplitude of the unstable input region of the inverter (e.g., the considered unstable region is $V_{IH} - V_{IL} + 2V_T$, where the thermal voltage also increases when increasing the temperature). Indeed, the percentage of unstable bits increases up to 1.36% of which only the 0.2% of flipped bits. However, I would stress that the percentage of flipped bits under temperature variations is underestimated by the circuit simulator, for the reason explained in the previous sub-chapter.

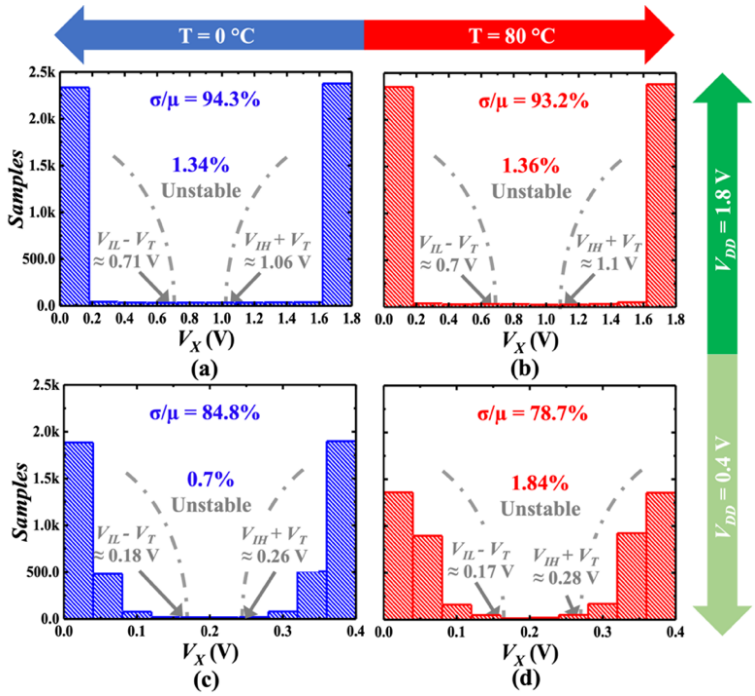


Fig. 3.32. Statistical distribution of the voltage V_X of the 4T bitcell core at different VT corners from 5k-run Monte Carlo simulations: (a) V_{DD} = 1.8 V and T = 0 °C, (b) V_{DD} = 1.8 V and T = 100 °C, (c) V_{DD} = 0.4 V and T = 0 °C, and (d) V_{DD} = 0.4 V and T =100 °C.

Then, Fig. 3.32(a)–(d) shows the V_X distributions (5k-run Monte Carlo simulations) at four different VT corners, along with the estimation of unstable bits. From this figure, the worst case is associated with the low V_{DD}, high temperature (0.4 V, 80 ◦C) corner where the percentage of unstable bits is ~1.8%. Fig. 3.33(a) and (b) report the percentage of the unstable bits under voltage and temperature variations, respectively, across different process corners.

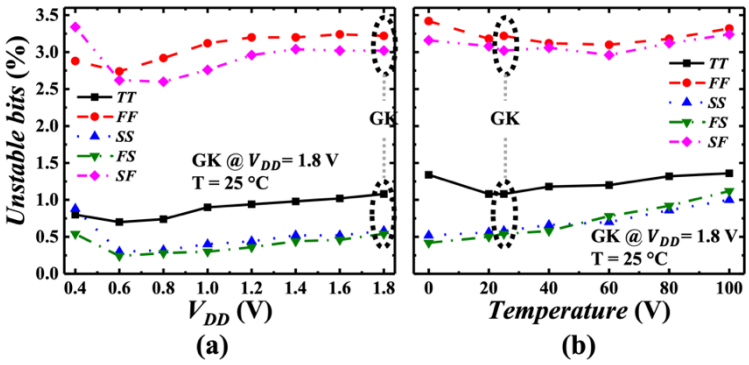


Fig.3.33. Percentage of total unstable bits for the 4T-based solution across different process corners under (a) voltage variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V.

The overall trend showed in Fig. 3.33(a) is similar to that reported at TT corner. Indeed, the percentage of unstable bits decreases when reducing the V_{DD} compared to the value at GK conditions due to narrower unstable region of the inverter. However, the worst corners are FF and SF due to faster PMOS thus resulting in a lower V_X spread for the reasons explained above. In particular, these two corners show a native percentage of unstable bits of 3.22% and 3.02% at GK conditions and 2.88% and 3.44% at V_{DD}= 0.4 V, respectively. On the other hand, the best corners are SS and FS in which slower PMOS ensures higher V_X deviation from the mid-supply point. Indeed, the percentage of unstable bits for SS and FS corners are 0.58% and 0.54% at GK conditions and 0.88% and 0.54% at V_{DD}= 0.4 V, respectively. The observed worsening at low voltage is strictly related to the inverter behavior under the same conditions. Fig. 3.33(b) report the percentage of unstable bits under temperature variations at different process corners. The showed trend is quite similar to that reported at TT conditions. The overall instability increases when increasing the temperature due to both the reduction of the strength difference and the increase of the unstable input region. Like the behavior under voltage variations, the worst corners are FF and SF whose percentage of unstable bits increase up to 3.32% and 3.24%, respectively when increasing the temperature up to 100 °C. On the other hand, the best corners are SS and FS where the unstable bits increase up to 1% and 1.12% with the same temperature variation. Concerning the power consumption, Fig. 3.34 shows the simulated absorbed current by the 4T core and by the bitcell under voltage and temperature variations at different process corners. Fig. 3.34(a) and (b) show the absorbed current by the core circuit (i.e., I_{4T}) under voltage and temperature variations, respectively. From Fig. 3.34(a) the I_{4T} shows a nearly independent trend under voltage variations due to the shielding effect provided by M3 and M4. Indeed, at TT corner the current (dissipated power) decreases from 35.90 pA (64.62 pW) to 34.96 pA (13.98 pW) when decreasing the V_{DD} down to 0.4 V. On the other hand, from Fig. 3.34(b) I_{4T} varies exponentially under temperature variations. Indeed, the current (dissipated power) increases from 35.90 pA (64.62 pW) to 772.9 pA (1.39 nW) when increasing the temperature up to 100 °C. Fig. 3.34(c) and (d) report the absorbed current by the bitcell (I_{DD}) under voltage and temperature variations,

respectively. From Fig. 3.34(c) the I_{DD} decreases exponentially when reducing the V_{DD}. Indeed, the current (dissipated power) decreases from 46.58 nA (83.84 nW) to 334.50 pA (133.80 pW) when decreasing the V_{DD} down to 0.4 V. On the other hand, the same current shows a nearly independent trend under temperature variations as reported in Fig. 3.34(d). Indeed, the I_{DD} (P_{DD}) increases from 46.58 nA (83.84 nW) up to 64.91 nA (116.84 nW) when increasing the temperature up to 100 °C.

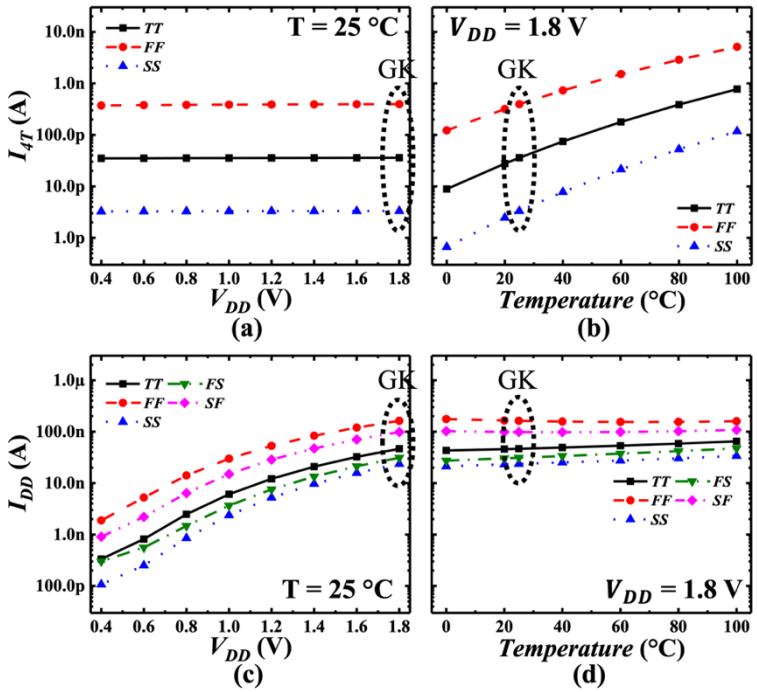


Fig. 3.34. Simulated absorbed current by the 4T-core across different process corners from 5k-run Monte Carlo simulations under (a) voltage variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V. Simulated absorbed currend across different process corners by the bitcell under (c) voltage variations at T= 25°C and (d) temperature variations at V_{DD}= 1.8 V.

Like the 2T-core, the worst corner is the FF where the higher PMOS conductivity increases the I_{4T} (P_{4T}) and the I_{DD}(P_{DD}) up to 395.5 pA (711.9 pW) and 162.3 nA (292.14 nW), respectively, at GK conditions. On the other hand, the best corner is the SS where the lower PMOS conductivity reduces the I_{4T} (P_{4T}) and the I_{DD}(P_{DD}) down to 3.33 pA (5.99 pW) and 23.65 nA (42.57 nW), respectively, at GK conditions. However, the overall improvement over the 2T-core solution relies on the higher V_X spread which dramatically reduce the absorbed current by the output inverter. Indeed, at GK conditions the power consumption decreases from 1.40 µW down to 83.84 nW (e.g., around two decades) when passing from the 2T-core to the 4T-core solution.

3.3.5 Measurements of the 4T-core based array

To prove the effectiveness of the proposed PUF concept, an 8 × 32 (i.e., 256 bit) bitcell array was implemented in a 180-nm test chip using transistor size and flavor as in Fig. 3.26. Fig. 3.35 illustrates the architecture of the PUF array, it was organized in four blocks, each including an 8 × 8 bitcell subarray and 3-to-8 decoder. The latter is used to select one row within the four 8 × 8 subarrays depending on its input signal ADDR_ROW. The row selection is enabled by inserting

an NMOS pass transistor proper sized for ensuring a correct operation across all the considered voltages (i.e., L= 0.9 µm and W= 0.44 µm) at the output node of each PUF bitcell, as showed at top left side of Fig. 3.35. The pass transistor also ensures a correct bitcell isolation along the output bitcells belonging to the same column.

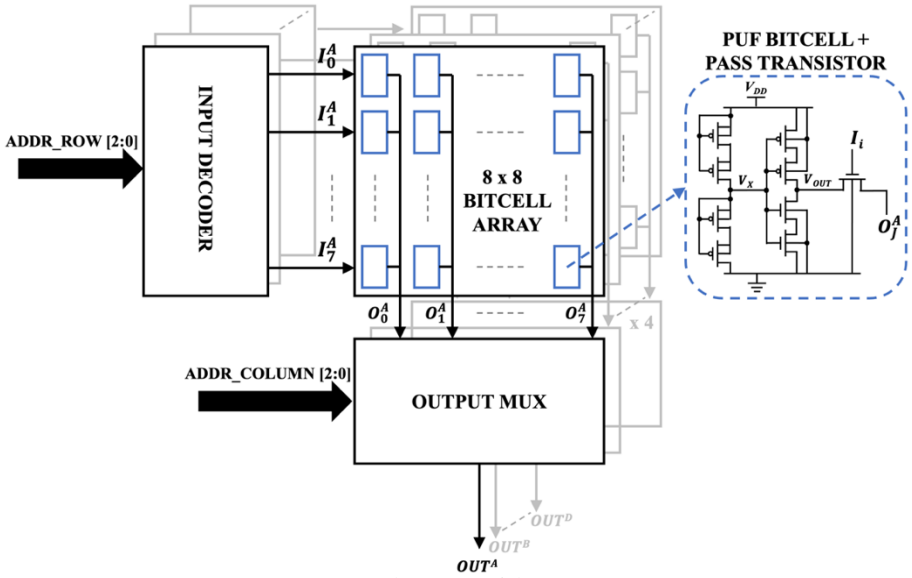


Fig. 3.35. Architecture of the PUF array.

Finally, the column is selected by an output multiplexer, which receives three signals ADDR_COL and provides four output bits (i.e., one bit for each 8 × 8 subarray). Fig. 3.36 illustrates the test chip.

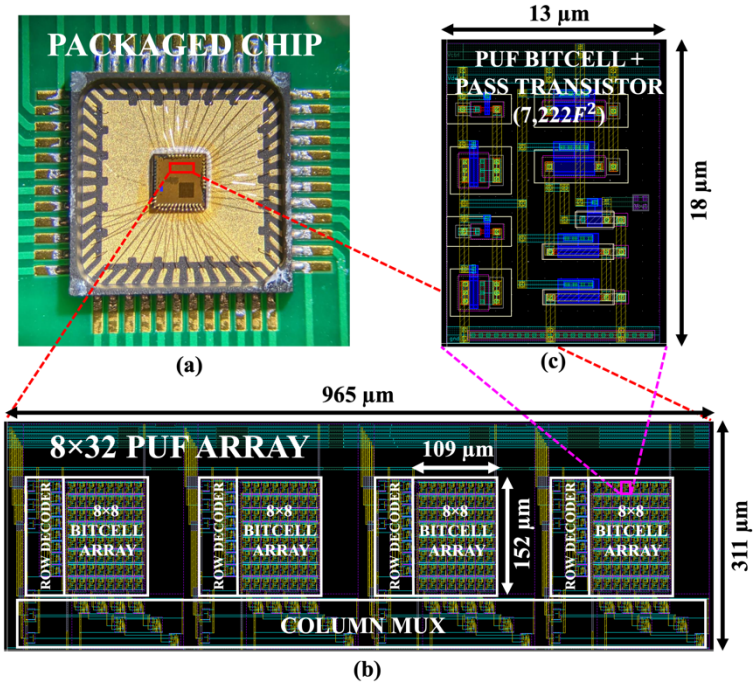


Fig. 3.36. (a) Photograph of the packaged test chip and layouts of (b) 8 × 32 PUF array, and (c) PUF bitcell area including pass transistor.

In particular, Fig. 3.36(a) shows the photograph of the fabricated (packaged) 180-nm test chip, whereas Fig. 3.36(b)-(c) shows the corresponding layouts of implemented circuit blocks. From Fig. 3.36(b), the whole 8 × 32 PUF array, including readout circuitry, occupies a silicon area of ~300,000 μm^2 (311 µm × 965 µm). Each 8 × 8 bitcell subarray entails an area occupation of 16,568 μm^2 (152 µm × 109 µm), corresponding to ~259-μm^2 area per bitcell. This is quite in agreement with the 234-μm^2 (18 µm × 13 µm corresponding to ~7,222F^2) bitcell area shown in Fig. 3.36(c), which includes the 4T-core, the output inverter, and the pass transistor. Measurements on the 8 × 32 PUF array were carried out across seven packaged dice (i.e., 1792 bits) using a custom PCB assisted by a Digilent Nexys 4 Artix-7 FPGA Trainer Board, a Rohde & Schwarz NGL202 power supply, and a Temptronic ThermoSpot DCP-101 system. Fig. 3.37(a)-(d) show the measurement results of the 8 × 32 PUF array at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C) across seven dice.

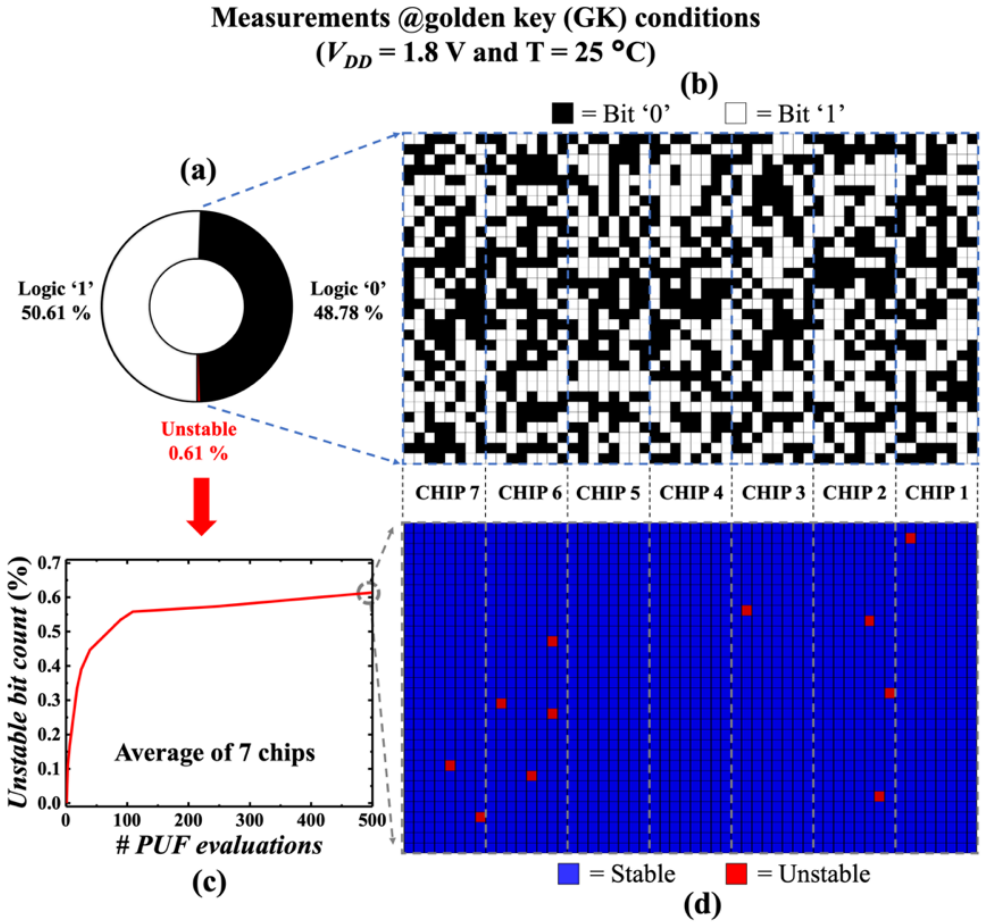


Fig. 3.37. Measurements of the 8 × 32 PUF array at GK conditions (V_{DD}= 1.8 V and T= 25 °C) across seven dice: (a) percentage of bit '0', bit '1', and unstable bits; (b) logical speckle diagram; (c) percentage of unstable bits versus number of evaluations; and (d) unstable bit mask at 500 evaluations.

Fig. 3.37(a) and (b) show, respectively, the corresponding breakdown among '0', '1', and unstable bits and the logical speckle diagram. The percentage of unstable bits at GK conditions was extracted from stability testing by performing 500 evaluations while marking bits that change at least once under different evaluations due to on-chip noise [4]. Fig. 3.37(c) reports the percentage of unstable bits versus the number of evaluations while Fig. 3.37(d) illustrates the unstable bit mask at 500 evaluations. Raw measurements reveal a native bit instability of 0.61% at GK

conditions, while the percentage of bit '0' and '1' are, respectively, 48.78% and 50.61%. Moreover, from Fig. 3.37(d) CHIP 4 and CHIP 5 are the most stable (i.e., no bit flip under different evaluations) while CHIP 6 is the least stable with 1.56% of unstable bits. The impact of voltage and temperature variations on the PUF stability is separately evaluated in Fig. 3.38(a) and (b), respectively. The former shows the percentage of unstable bits under V_{DD} variations (0.4−1.8 V range) at T= 25 °C, whereas the latter reports the bit instability across temperature variations (10−80 °C range) at V_{DD}= 1.8 V. In both figures, the total unstable bits (averaged over seven dice) are given by the sum of two different contributions. The first one is due only to on-chip noise (i.e., related to unstable noisy bits that change at least once under different evaluations for a given voltage and temperature). The second contribution refers to bits that stably flip under environmental variations (i.e., voltage and/or temperature variations) compared to the nominal condition (i.e., GK conditions) while discarding the unstable bits due to on-chip noisy.

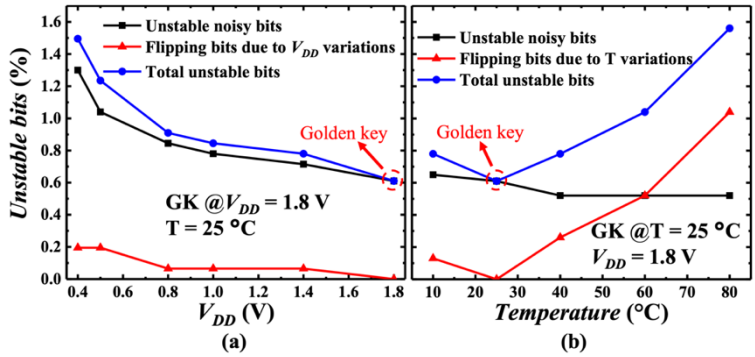


Fig. 3.38. Percentage of unstable bits (averaged over seven dice) under (a) V_{DD} variations at T= 25 °C and (b) temperature variations at V_{DD}= 1.8 V.

From Fig. 3.38(a) the percentage of total unstable bits increase from 0.61 up to 1.495% when decreasing the V_{DD} down to 0.4 V. Such increase of the overall instability is mainly due to the increase of noisy bits. Indeed, under 1.8−0.4 V voltage variations the noisy and flipped bits increase up to 1.3% and 0.195%, respectively. As predicted by the simulations, the impact of flipped bits is very low under voltage variations thus proving the effectiveness of the shielding effect provided by M3-M4. However, the increase of the noisy bits can be probably ascribed by the non-full swing V_X voltages (i.e., when the mismatch is very small) at the output of the conversion stage only one conversion stage along with the use of a NMOS pass transistor for isolating the bitcell along the bitcells belonging to the same column. In particular, in the case of very small mismatch the gain provided by the 4T-core is not high enough for guarantying full swing output voltage as well as that of the output inverter. Moreover, when decreasing the V_{DD} the off currents of the pass transistors belonging to the same columns could affect the output voltage of the accessed bitcell. This probably results in higher noise sensitivity at the input of the output MUX. From Fig. 3.38(b), temperature variations result in an increase of the unstable bits. Indeed, the percentage of total unstable bits increase from 0.61% up to 1.56% when increasing the temperature up to 80 °C. However, unlike voltage variations the overall instability under temperature variations is dominated by flipped bits, especially for temperature above 60 °C. Indeed, the noisy bits remains quite constant across the considered temperature range (e.g., between 0.52% and 0.65%) while the flipped bits increase up to 1.04% when increasing the temperature up to 80 °C. This is mainly due to the mismatch in terms of V_{TH} temperature

coefficient which flips the $\Delta V_{TH1,2}$ polarity under temperature variations. Indeed, according to Fig. 3.4 and Fig. 3.8, Fig. 3.39 illustrates an example of M1 and M2 threshold voltage trends under temperature variations at different mismatch conditions. In all cases the mismatch (i.e., $V_{TH0,1}$ - $V_{TH0,2}$) at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C) is of 1 mV. In Fig. 3.39(a) the two threshold voltage temperature coefficients are equal (i.e., $k_{T,1} = k_{T,2}$) and then no flip occurs across the considered temperature range. On the other hand, in Fig. 3.39(b) and (c) the two coefficients differ from each other.

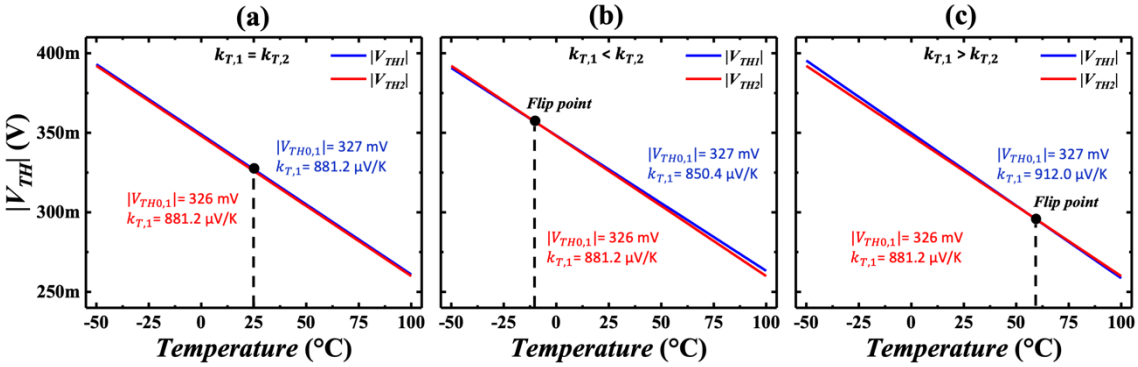


Fig. 3.39. Example of M1 and M2 threshold voltages trend under temperature variations when (a) $k_{T,1} = k_{T,2}$, (b) $k_{T,1} < k_{T,2}$, and (c) $k_{T,1} > k_{T,2}$.

In particular, from Fig. 3.39(b) the temperature coefficient of M1 is lower than that of M2 and then the flip occurs for temperature lower than the nominal one (i.e., T= 25 °C). On the other hand, in Fig. 3.39(c) the temperature coefficient of M1 is higher than that of M2 and then the flip occurs for temperature higher than the nominal one. The addition of M3-M4 counteracts the effect of DIBL coefficient mismatch between M1 and M2 limiting the voltage drops across them. On the other hand, preserving the mismatch polarity under temperature variations still represents a crucial issue despite the use of two nominally identical sub-circuit with unique transistor flavor reduce the variability of the temperature coefficient. Anyway, data of Fig. 3.38(a) and (b) show that the stability degrades when decreasing the voltage and/or varying the temperature. This suggests that the worst-case stability characterization requires testing at the low V_{DD} and high temperature corner (i.e., for the considered temperature range the worst-case corner is at high temperature). Stability measurements were performed at different VT corners for two of the seven chips (i.e., the chip 1 and 2 reported in Fig. 3.37(b) and (d)). These two chips show an average native instability of 0.8% at GK conditions, which is the same obtained at V_{DD}= 1.8 V and T= 10 °C corner. The instability increases up to ~1.4% and ~1.9%, respectively, at V_{DD}= 0.4 V and T= 10 °C and V_{DD}= 1.8 V and T= 80 °C corners. As predicted by the simulations the worst case is associated with V_{DD}= 0.4 V and T= 80 °C where the instability reaches ~2.3%. Fig. 3.40(a) and (b) report the bit error rate (BER) measured across voltages and temperatures while considering 32-bit PUF output words. In both figures, the BER shows a trend similar to the total instability showed in Fig. 3.38. Indeed, from Fig. 3.40(a) the BER increases from 0.13% (at GK conditions) up to 0.87% when decreasing the voltage down to 0.4 V. Similarly, from Fig. 3.40(b) the BER deviates from its nominal value under voltage variations. Indeed, it increases from 0.13% up to 1.13% when increasing the temperature up to 80 °C.

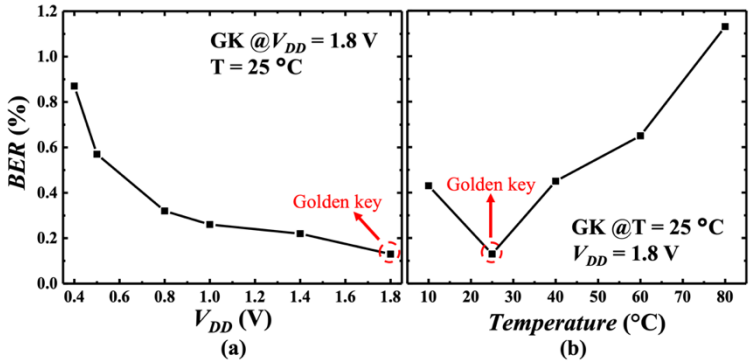


Fig. 3.40. BER under (a) V_{DD} variations at T= 25 °C and (b) temperature variations at V_{DD} = 1.8 V. Data are averaged over seven dice considering 32-bit PUF words.

Fig. 3.41(a)-(d) report some PUF metrics such as uniqueness, reproducibility, and randomness [23], estimated across seven dice for 32-bit PUF words.

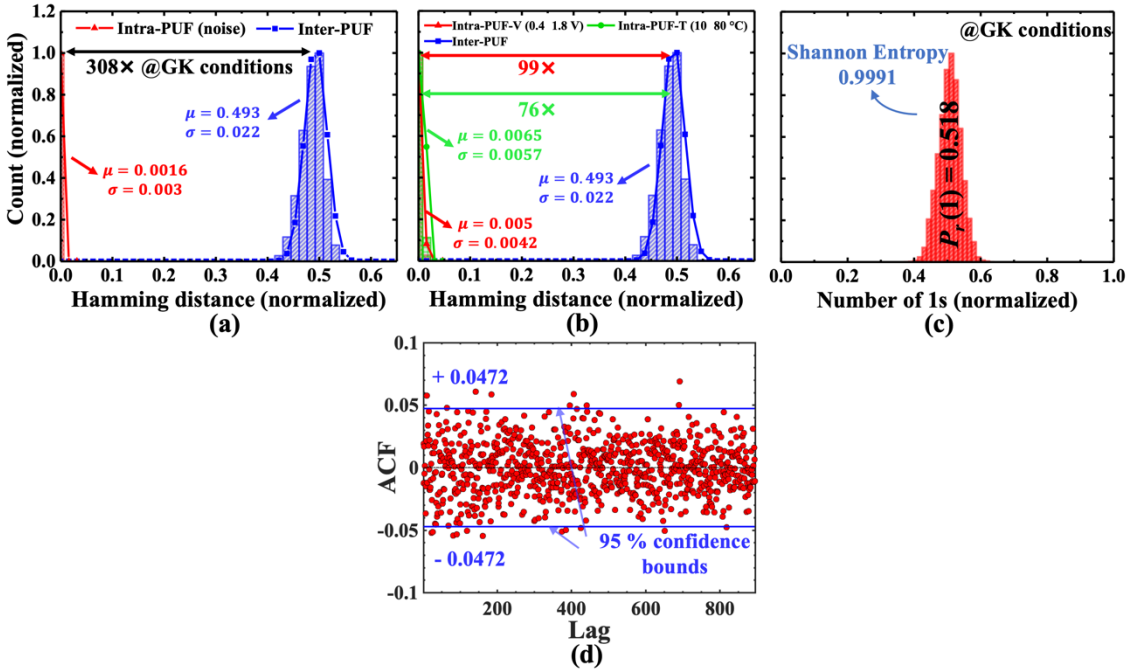


Fig. 3.41. (a) Normalized inter-PUF and intra-PUF HD at GK conditions (i.e., due only to on-chip noise), (b) normalized intra-PUF HD under voltage and temperature variations, (c) normalized number of bit "1" at GK conditions. Data are evaluated across seven dice considering 32-bit PUF words, and (d) spatial autocorrelation function (ACF).

The uniqueness (i.e., the ability to generate unique identification across different dice for the same input challenge) was evaluated through the normalized (i.e., to the length of the PUF response) inter-PUF (i.e., considering the responses for the same challenge of PUF instances implemented in different chips) Hamming distance (HD). From Fig. 3.41(a), the normalized inter-PUF HD shows a mean value of 0.493, which is quite close to the ideal value of 0.5. The reproducibility (i.e., the the ability to generate a consistent response regardless of on-chip noise effect and variations in the environmental conditions) was estimated through the normalized intra-PUF HD (i.e., considering the responses for the same challenge of the same PUF instance but evaluated under noisy or different environmental conditions) measured at GK (i.e., due only to on-chip noise) and under different environmental conditions (i.e., under voltage and temperature

variations) as reported in Fig. 3.41(a) and (b). From Fig. 3.41(a), the intra-PUF HD under noisy conditions shows a mean value of 0.0016 which is quite close to the ideal value of zero. This leads to an identifiability (i.e., the ability of showing a distinguishable behavior compared to other instances under noisy or different environmental conditions quantified by the ratio between inter and intra HD) of 308× at GK conditions. From Fig. 3.41(b), the normalized intra-PUF shows a mean value under voltage (0.4−1.8 V) and temperature (10−80 °C) variations of 0.0057 and 0.0065, respectively. This translates in a good identifiability of 99× and 76×, respectively, thus highlighting the ability of the PUF instance of being distinguishable even under environmental variations. The randomness (i.e., the PUF unpredictability) was estimated by calculating the probability of generating a bit "1" (i.e., $P_r(1)$) in the 32-bit PUF word at GK conditions as well as by evaluating the spatial autocorrelation function (ACF) on the spatial distribution from different dice, as shown in Fig. 3.41(c) and (d). From Fig. 3.41(c) the mean $P_r(1)$ of 0.518 results in a Shannon Entropy of 0.9991, these values are quite close to the ideal values of 0.5 and 1, respectively. From Fig. 3.41(d), the spatial ACF at 95% confidence bounds is 0.0472 which is very close to the ideal value of zero, thus proving good rejection of layout-dependent variations. The randomness was more rigorously assessed by performing statistical NIST test [5], whose results are reported in Table I. For each test (those requiring stream length n > 3040 were neglected [4]), the *p-value* averaged over seven dice was evaluated. A *p-value* greater than 0.01 is desired to consider an arbitrary source of information random with 99% confidence, and higher values indicate a higher confidence about the source randomness

TABLE I. NIST TEST RESULTS (AVERAGE OVER 7 DICE)

NIST test	Stream length (n)	# Runs	Average p-value	PASS (%)	PASS (?)
Frequency (Freq)	256	7	0.6279	100	YES
Block frequency (BF)	256	7	0.6219	100	YES
Runs	256	7	0.6267	100	YES
Longest runs of ones (LRO)	256	7	0.4914	100	YES
FFT	256	7	0.3950	100	YES
Nonoverlapping template (NOT)	256 (m = 4)	7	0.5592	100	YES
Serial	256 (m = 4)	7	0.6276	100	YES
Approximate entropy (AppEn)	256 (m = 2)	7	0.6413	100	YES
Cumulative sums (CumSum)	256	7	0.6396	100	YES

From Table I, all tested dice pass all implemented NIST tests with averaged *p-values* well above 0.01, thus indicating good confidence about the source randomness. We also evaluated the static power consumption in the PUF array, while excluding the contribution of readout circuitry in Fig. 3.35. In this regard, Fig. 3.42(a) and (b) plots the measured supply current per bitcell versus the supply voltage at T = 25 °C for six of seven dice and averaged over seven dice, respectively. In particular, as shown in Fig. 3.26(a) the supply current is given by the sum of two contributions. The first one is the absorbed current by the 4T-core (i.e., I_{4T}) while the second contribution is that absorbed by the output inverter (i.e., I_{INV}). The latter is dominant regardless of the operating voltage and is function of the V_X voltage (i.e., the output voltage of the 4T-core). This explains the increasing trend of the supply current when increasing the V_{DD}.

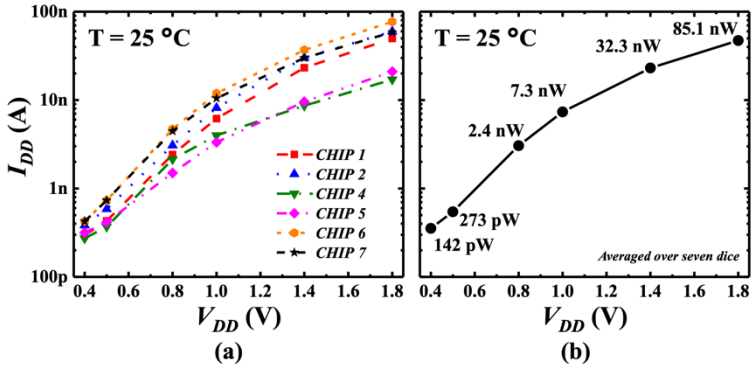


Fig. 3.42. Measured supply current (I_{DD}) per bitcell versus V_{DD} at T= 25 °C for (a) six of seven dice, and (b) averaged over seven dice with relative static power.

Indeed, the absorbed current by the 4T-core shows a nearly constant trend under voltage variation as predicted and verified by simulation and measurement, respectively, showed in both Fig. 3.24 and Fig. 3.25. From Fig. 3.42(a) the lowest supply current per bitcell corresponds to CHIP 4 and CHIP 5 which also are the most stable at while the highest supply current corresponds to CHIP 6 which is the least stable GK conditions. This can be ascribed by the fact that higher absorbed current refers to a higher percentage of V_X samples that fall in the unstable input region of the inverter thus increasing the supply current of the output inverter as well as resulting potentially unstable at the output. On the other hand, from Fig. 3.42(b) the average supply current (power consumption) per bitcell at GK conditions is 47.23 nA (85.1 nW) which decrease down to 353.85 pA (142 pW) when decreasing the V_{DD} down to 0.4 V. The measured supply current, and hence the power consumption, are very similar to the simulated data showed in Fig. 3.34, thus proving the goodness of the simulations reported above.

3.3.6 Comparison with prior works

Finally, Table II summarizes the main metrics evaluated for the implemented PUF solution and the comparison with relevant and recent prior art CMOS PUF designs. The table includes only measured data achieved without applying any post-silicon stability-enhancement techniques (if not differently specified) for a fair comparison. From Table **Errore. L'origine riferimento non è stata trovata.**, the achieved percentage of native unstable bits of 0.61% is 2.7-44× lower than prior art, whereas the native BER of 0.13% is equal to [46] and 1.6-44× lower than other designs. Despite the lowest minimum operating voltage of only 0.4 V and the wider voltage range considered in this work, the voltage dependence of the bit instability (0.63%/V) is close to the value reported in [47] and 2.1-7.4× lower than other PUFs. At the same time, the dependence of the bit instability on temperature (0.016%/°C) is similar to [50], [46], [47], and 2.8-6.9× lower than [34], [49], [58]. Concerning the statistical metrics, the absolute deviation of the normalized inter-PUF HD from the ideal value is 0.007, i.e., worse than [34], [49], [58]-[37], similar to [29], and better than [30], [43], [50]. Also, the normalized intra-PUF HD of 0.0016 is nearly the same as [34], 2.3× higher than [46], and 1.9-4.3× lower than other designs. The proposed solution also shows the third best value of identifiability (308×), which is lower than [34] and [46] respectively by 1.1× and 2.3×, and 1.9-22× higher than other implementations.

TABLE II. PUF METRICS SUMMARY AND COMPARISON WITH STATE-OF-THE-ART CMOS PUF DESIGNS
(ONLY MEASURED DATA W/O STABILITY-ENHANCEMENT TECHNIQUES IF NOT DIFFERENTLY SPECIFIED, BEST PERFORMANCE IN BOLD)

	This work [13]	**JSSC 2018 [34]**	**JSSC 2020 [29]**	**JSSC 2021 [30]**	**JSSC 2017 [43]**	**JSSC 2016 [49]**	**JSSC 2018 [50]**	**JSSC 2016 [58]**	**ISSCC 2017 [46]**	**JSSC 2020 [47]**	**ISSCC 2021 [37]**
tech. [nm]	180	65	130	130	14	65	40	65	180	65	180
bitcell type	4T voltage divider	SRAM	SRAM	SRAM	metastable	current mirror	cascode current mirror	PTAT voltage generator	2T amplifier (DNW)	3T amplifier	leakage based (footed)
bitcell area [F^2]	7,222	600	**373**	497	9,388	6,000	3,644	727	553	562	890
native unstable bits [%] (# of evaluations)	**0.61** (500)	5.39 (500)	2.14 (2000)	2.71 (1000)	26.8 (5000)	1.73 (400)	2.55 (500)	6.54 (500)	1.67 (2000)	2.95 (2000)	5.62 (1000)
native BER [%]	**0.13**	2.16	0.21	0.29	5.76	N/A	0.81	N/A	**0.13**	0.3	0.69
mean $P_r(1)$ (|deviation| from 0.5[a])	0.518 (0.018)	N/A	N/A	N/A	N/A	**0.502 (0.002)**	0.52 (0.02)	N/A	N/A	N/A	N/A
V_{DD} range [V]	0.4–1.8	1.0	0.8–1.4	0.5–0.7	0.55–0.75	0.6–1.0	0.8–1.0	0.6–1.2	0.8–1.8	0.7–1.4	1.2–1.8
bit instability voltage dependence [%/V]	0.63	N/A	N/A	N/A	N/A	4.68	3.6	1.3	2[b]	**0.57**	N/A
T range [°C]	10–80	-15–85	-40–120	-40–120	25–110	25–85	-40–125	0–80	-40–120	-55–125	0–80
bit instability temp. dependence [%/°C]	0.016	0.11	N/A	N/A	N/A	0.047	0.015	0.044	0.02[b]	**0.012**	N/A
mean normalized inter-PUF HD (|deviation| from 0.5[a])	0.493 (0.007)	0.502 (0.002)	0.4923 (0.0077)	0.4873 (0.0127)	0.486 (0.014)	0.5014 (0.0014)	0.4907 (0.0093)	0.5001 (0.0001)	0.4989 (0.0011)	0.4998 (0.0002)	**0.50001 (0.00001)**
mean normalized intra-PUF HD (noise)	0.0016	0.00151	0.003	0.0041	0.034	0.0034	0.0049	0.0057	**0.0007**	0.0047	0.00685
inter/intra HD ratio (identifiability)	308×	332×	164×	119×	14×	149×	102×	88×	**713×**	106×	73×
spatial ACF @95% confidence	0.0472	N/A	0.0228	0.0334	N/A	0.0363	**0.00735**	0.0188	0.0167	0.01385	0.022
static power per bitcell [W]	85n (1.8V, 25°C)	N/A	N/A	N/A	3u (0.65V, 70°C)	N/A	8n[c] (0.9V, 25°C)	N/A	**26p** (1.2V, 25°C)	N/A	N/A

[a] ideal value [b] after temporal majority voting (TMV) [c] with power gating

Note that above performance in terms of stability and reproducibility are achieved at the cost of larger area occupation with respect to previous PUF solutions. Indeed, the ~7,200-F^2 bitcell area is 1.2-19.4× higher than other designs in Table II, with the exception of [43]. From Table II, the proposed solution also shows a penalty in terms of spatial ACF (0.0472 at 95% confidence), which is 1.3-6.4× higher than prior works. Moreover, the measured static power per bitcell of 85 nW (1.8 V and 25 °C) is in the same order of magnitude with respect to that achieved in [50] with power gating, ~35× lower than [43], and higher than [46] by more than one order of magnitude.

3.4 Area-Stability trade-off

The results showed above were obtained at the cost of a larger area compared to several other prior works. The adoption of a sub-threshold voltage divider between two nominally identical sub-circuits guarantees an adequate randomness regardless of the process variations while also ensuring an inherently reliability under voltage and temperature variations. Indeed, the proposed solution showed low native instability as well as a correct operation under voltage and temperature variations within 0.4−1.8 V and 10−80 °C, respectively. Moreover, the lower voltage bound is dictated by the output inverter. Indeed, simulation and measurement results of the 4T-core, showed in Fig. 3.24, prove its ability of operating at very low voltages (i.e., with V_{DD} down to 0.1 V). However, the above considerations rely on the first crucial assumption of having two nominally identical sub-circuits (i.e., TC ≡ BC) which constraints the minimum area. For example, using a NMOS-based bottom circuit instead of the proposed PMOS version helps reducing the bitcell footprint breaking, however, the inherent symmetry of the circuit. This results in a higher sensitivity to PVT variations. Indeed, the systematic variations at different process corners causes that the randomness is not respected, due to the high gain provided by the core circuit, which requires external circuits for balancing the PMOS and NMOS strength under process variations.

This section provides different circuital variants, with the aim of reducing the bitcell footprint without degrading such fundamental PUF metrics.

3.4.1 Area reduction of the output inverter

First, the area efficiency can be improved reducing the area of the output inverter. Indeed, using a minimum inverter (i.e., minimum sizing for ensuring balanced PMOS and NMOS strength) instead of that used in the proposed solution helps reducing the area overhead at the cost of a lower gain (and hence larger unstable input region), higher power consumption, and higher variability. Fig. 3.43(a) and (b) illustrate the schematic of the high-gain 4T inverter and low-area 4T inverter, respectively. In particular, the inverter showed in Fig. 3.43(a) is the same of that reported in the previous sub-section. On the other hand, the inverter proposed here is minimum sized where M1 and M4 are used for enabling power gating for turning off the unused cells.

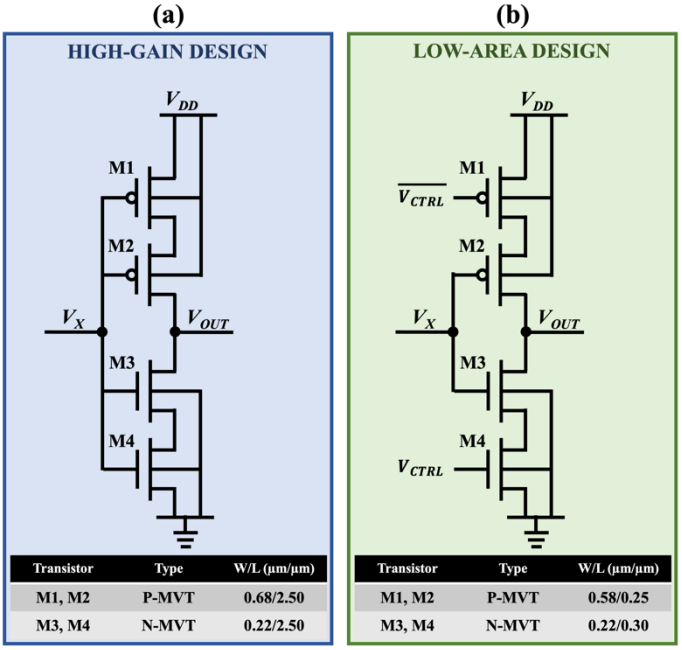


Fig. 3.43. Schematic of (a) high-gain 4T inverter design, and (b) low-area 4T inverter design.

Fig. 3.44(a)-(d) report the static behavior of the two inverters (i.e., high-gain inverter, which is the adopted inverter for the previous proposed designs, and low-area inverter, which is the proposed area-efficient design) under voltage and temperature variations at TT corner. Fig. 3.44(a) and (b) show the amplitude of the input unstable region (V_{IH} - V_{IL}) as function of voltage and temperature variations, respectively. From Fig. 3.44(a) the low-area design shows larger unstable input region under voltage variations, especially for higher V_{DD} thus potentially resulting in a higher percentage of unstable bits. On the other hand, from Fig. 3.44(b) both designs exhibit a slight increase under temperature variations. Fig. 3.44(c) and (d) show the behavior of the input logic threshold (V_M) for both designs under voltage and temperature variations, respectively. From Fig.3.44(c) the logic threshold of both designs follows the mid-supply point (V_{DD}/2) when decreasing V_{DD} down to 0.4 V. However, at very low voltage operation the high-gain design deviates further away from V_{DD}/2 than the low-area design. Finally.

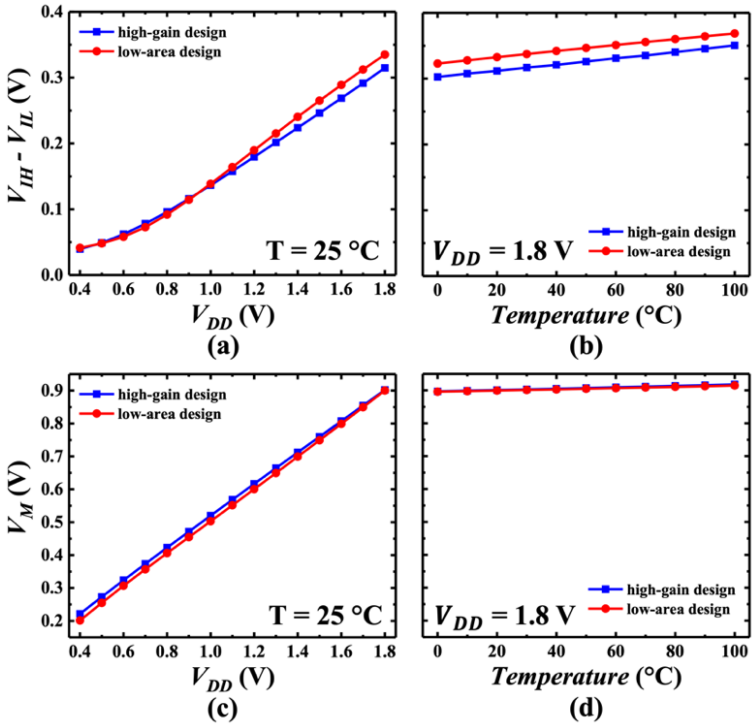


Fig. 3.44. Effect of voltage and temperature variations on the static parameters of the output inverter for both high-gain design and low-area design. Effect of (a) voltage with T = 25 °C and (b) temperature with V_{DD} = 1.8 V variations on the difference between minimum high- and maximum low-input voltages (V_{IH}-V_{IL}). Effect of (c) voltage with T = 25 °C and (d) temperature with V_{DD} = 1.8 V variations on the input logic-threshold (V_M).

From Fig. 3.44(d) the logic threshold of both inverters exhibits the same slight increasing trend when increasing the temperature up to 100 °C. Concerning the variability, Fig. 3.45(a)-(d) show the statistical distribution of the amplitude of the unstable region (i.e., V_{IH} - V_{IL}) and the logic threshold (i.e., V_M) at V_{DD}= 1.8 V and 0.4 V with T= 25 °C. Data came from 5k-run Monte Carlo simulations at TT corner. Fig. 3.45(a) and (c) report the statistical distributions of the input unstable region at V_{DD}= 1.8 V and V_{DD}= 0.4 V, respectively, with T= 25 °C. From these distributions, the low-inverter design shows an average unstable region larger than the high-gain design of 20.8 mV and 2 mV at 1.8 V and 0.4 V, respectively. Moreover, the variability of the low-area design (i.e., the ratio between mean value and standard deviation) increases from 0.75% up to 1.08% when decreasing the V_{DD} down to 0.4 V while that of the high-gain design increases from 0.27% to 0.66%, respectively. Fig. 3.45(b) and (d) illustrate the statistical distributions of the input logic threshold at V_{DD}= 1.8 V and V_{DD}= 0.4 V, respectively, with T= 25 °C. From Fig. 3.45(b) the mean value is quite close to the ideal one (i.e., V_M= 0.9 V) while the variability is 0.57% for the low-area design and 0.17% for the high-gain design. Fig. 3.45(d) highlights that when decreasing the V_{DD} down to 0.4 V the variability of the low-area and high-gain designs increase up to 2.78% and 0.59%, respectively. On the other hand, the mean values are 201.4 mV and 221.6 mV, respectively, demonstrating that the pull-up and the pull-down network strengths at very low-voltages are better balanced in the low-area design than the high-gain design. Finally, Fig. 3.46 (a) and (b) show the short-circuit current ($I_{SC,INV}$) of the inverter (i.e., when $V_X = V_M$) for both low-area and high-gain designs under voltage and temperature variations, respectively.

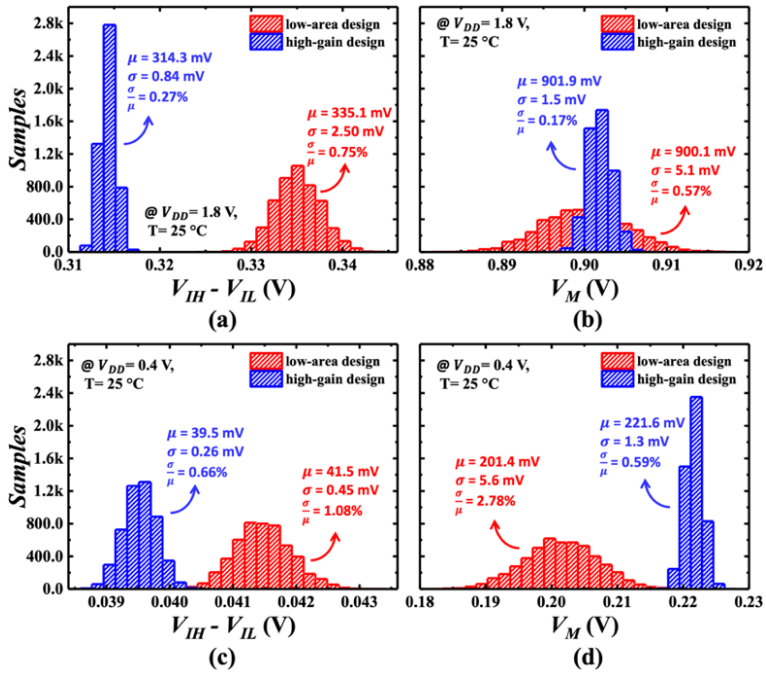


Fig. 3.45. Statistical distribution of the amplitude of the unstable region (i.e., V_{IH} - V_{IL}) and the input logic logic threshold (i.e., V_M) at (a) and (b) V_{DD}= 1.8 V and T= 25 °C, and (c) and (d) V_{DD}= 0.4 V and T= 25 °C, respectively, from 5k-run Monte Carlo simulations at TT corner.

From Fig. 3.46(a) for both designs the $I_{SC,INV}$ shows a decreasing trend when decreasing the V_{DD} down to 0.4 V. Indeed, when moving toward a low-area design, the $I_{SC,INV}$ (P_{SC}) increases from 2.62 µA (4.72 µW) up to 26.46 µA (47.63 µW) at nominal V_{DD} (i.e., 1.8 V) and from 11.89 nA (4.76 nW) up to 45.17 nA (18.07 nW) at 0.4 V.

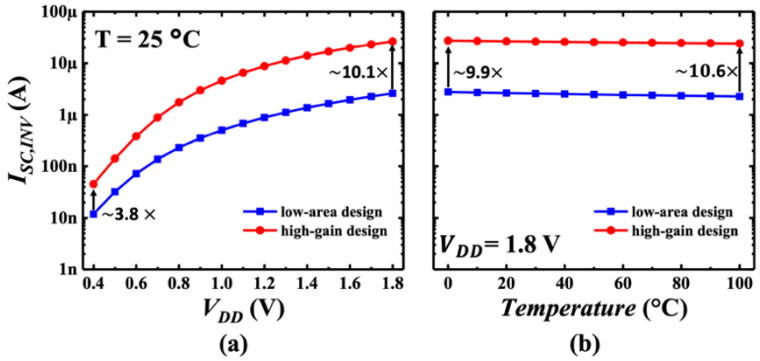


Fig. 3.46. Short-circuit current (i.e., when V_X = V_M) of the inverter under (a) V_{DD} variations at T= 25 °C and (b) Temperature variations at V_{DD}= 1.8 V.

On the other hand, from Fig. 3.46(b) the $I_{SC,INV}$ shows a decreasing trend when increasing the temperature. Indeed, using a low-area design instead of the high-gain version increases the $I_{SC,INV}$ (P_{SC}) from 2.78 µA (5 µW) up to 27.45 µA (49.41 µW) at T= 0°C and from 2.27 µA (4.09 µW) up to 24.09 µA (43.36 µW). Anyway, as shown in Fig. 3.43, M1 and M3 are driven by a control signal which allows enabling the power gating of the array. Indeed, during the readout phase only one row per time is enabled while powering off the unused cells. This approach dramatically reduces the power consumption of the array.

3.4.2 Area – Stability tradeoff in the 4T voltage divider

Another way for reducing the bitcell footprint is connecting the body terminal of the transistors in each sub-circuit to each other. Fig. 3.47(a)-(c) illustrate the bitcell design concept including the area-efficient inverter, described above, along with the schematics and layouts of the bitcells which rely on the 4T-core described in the previous sub-section, i.e., without including the body effect, and the more compact 4T-core, i.e., including the body effect.

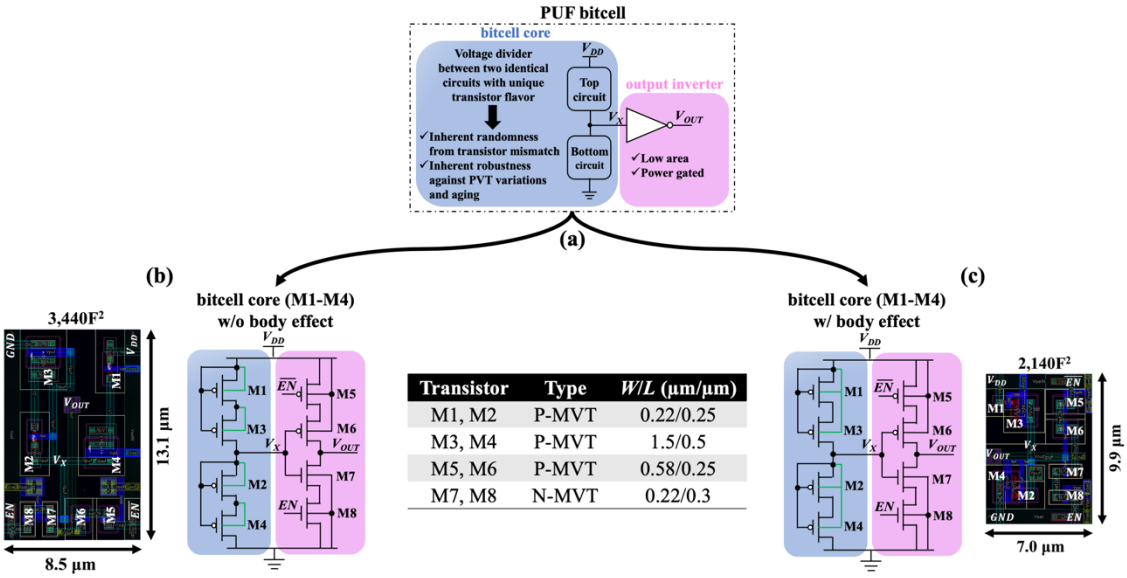


Fig. 3.47. (a) Bitcell design concept along with schematic and layout (b) without using the body effect and (c) using the body effect.

Fig. 3.47(a) provides the novel bitcell design concept. The main difference over that proposed in the previous sub-section is that here, the row selection (i.e., the isolation of a bitcell from the others belonging to the same column) is e enabled by the header and footer transistors in the output inverter. This approach also ensures of turning off the unused bitcells thus resulting in better energy efficiency which counteracts the power consumption increase due to the adoption of a minimum inverter instead of the high-gain design. Fig. 3.47(b) illustrates the schematic and layout of the 4T-core bitcell, pointing out that using an inverter with low-area inverter design, using the same sizing for the 4T-core, allows achieving 52.37% of area saving. Indeed, the bitcell footprint decreases from $7{,}222F^2$ down to $3{,}440F^2$. This area can be further reduced by including the body effect within the core circuit. Indeed, Fig. 3.47(c) provides the schematic and layout of the 4T-core based bitcell in which the body terminals of the transistors in each sub-circuit are connected to the source terminal of M1 and M2 for top and bottom circuit, respectively. This allows further achieving 37.79% of area saving. Indeed, the bitcell footprint decreases from $3{,}440F^2$ down to $2{,}140F^2$ when including the body effect within the core circuit due to the possibility of sharing the n-well of the two transistors in each sub-circuit. For better understanding the effect of including the body effect within the core circuit on the overall performance, in the following analytical derivation the terms in black are shared by the two circuits while the terms in green only refer to the core which includes the body effect. Using the same assumptions of the 4T core solution. The current equations (3.7) and (3.8) are the same for both core solutions, while the current of M3 and M4 are given by

$$I_{M3} = I_{0,3}\frac{W_3}{L_3}\exp\left(\frac{(1+\gamma_{B3,4})(V_1 - V_{DD}) + V_{TH0,3}(T) + \lambda_{D3,4}(V_1 - V_X)}{n_{3,4}V_T}\right) \quad (3.19)$$

$$I_{M4} = I_{0,4}\frac{W_4}{L_4}\exp\left(\frac{(1+\gamma_{B3,4})(V_2 - V_X) + V_{TH0,4}(T) + \lambda_{D3,4}(V_2)}{n_{3,4}V_T}\right) \quad (3.20)$$

where $\gamma_{B3,4}$ is the body biasing coefficient of M3 and M4. Following the same procedure developed for the 4T core and then equating (3.7) and (3.19) (i.e., the currents of M1 and M3), and (3.8) and (3.20) (i.e., the currents of M2 and M4), the voltages V_1 and V_2 are given by

$$V_1 = \frac{1}{n_{1,2} + n_{1,2}\lambda_{D3,4} + n_{3,4}\lambda_{D1,2} + n_{1,2}\gamma_{B3,4}}\Big[(n_{1,2} + n_{3,4}\lambda_{D1,2} + n_{1,2}\gamma_{B3,4})V_{DD} + n_{3,4}V_{TH0,1}(T) - n_{1,2}V_{TH0,3}(T) + n_{1,2}\lambda_{D3,4}V_X + n_{1,2}n_{3,4}V_T\ln\left(\frac{I_{0,1}W_1L_3}{I_{0,3}W_3L_1}\right)\Big] \quad (3.21)$$

$$V_2 = \frac{1}{n_{1,2} + n_{1,2}\lambda_{D3,4} + n_{3,4}\lambda_{D1,2} + n_{1,2}\gamma_{B3,4}}\Big[n_{3,4}V_{TH0,2}(T) - n_{1,2}V_{TH0,4}(T) + (n_{1,2} + n_{3,4}\lambda_{D1,2} + n_{1,2}\gamma_{B3,4})V_X + n_{1,2}n_{3,4}V_T\ln\left(\frac{I_{0,2}W_2L_4}{I_{0,4}W_4L_2}\right)\Big] \quad (3.22)$$

Then, by substituting (3.21) and (3.22), respectively, in (3.7) and (3.8) and equating the resulting expressions, the equation for the voltage V_X is given by

$$V_X = \frac{V_{DD}}{2} + \frac{1}{2\lambda_{D3,4}}\left[V_{TH0,3} - V_{TH0,4} + n_{3,4}V_T ln\left(\frac{I_{0,3}W_3L_4}{I_{0,4}W_4L_3}\right)\right] + \frac{1+\gamma_{B3,4}+\lambda_{D3,4}}{2\lambda_{D1,2}\lambda_{D3,4}}\left[V_{TH0,1} - V_{TH0,2} + (k_{T,1} - k_{T,2})(T - T_{nom}) + n_{1,2}V_T ln\left(\frac{I_{0,1}W_1L_2}{I_{0,2}W_2L_1}\right)\right] \quad (3.23)$$

From (3.23), the gain over the M1-M2 mismatch is slightly higher when including the body effect within the core circuit. For better understanding the circuit behavior, equation (3.18) can be rewritten by following the same procedure as for the core circuit of Fig. 3.47(b).

$$V_{SD3} - V_{SD,4} \approx \frac{V_{TH0,4}(T) - V_{TH0,3}(T)}{\lambda_{D3,4}} + \frac{1+\gamma_{B3,4}}{\lambda_{D3,4}}[V_{SD1}(T) - V_{SD2}(T)] \quad (3.24)$$

This equation points out that the voltage drops across M1 and M2 act on both gate and body terminal thus increasing the overall gain over their mismatch. This point can be better appreciated in Fig. 3.48(a)-(c) which report the statistical analysis of the V_X voltage and $V_{SD1} - V_{SD2}$ as function of the M1-M2 mismatch along with the statistical distribution of the circuit gain for both solutions of Fig. 3.47(b) and (c). Fig. 3.48(a) illustrates the V_X sensitivity of both circuits to the M1-M2 mismatch, from which it is notable a very similar input-output characteristic. This is confirmed by Fig. 3.48(c) which shows the statistical distributions of the gain of circuits of Fig. 3.47(b) and (c). From this figure the new design slightly increases the circuit gain from 1.899k to 2.202k (i.e., ×1.16). Indeed, thanks to both reverse gate and body bias the new design requires lower $V_{SD1} - V_{SD2}$ for achieving even higher V_X spread than that of the circuit of Fig. 3.47(b) as shown in Fig. 3.48(b).

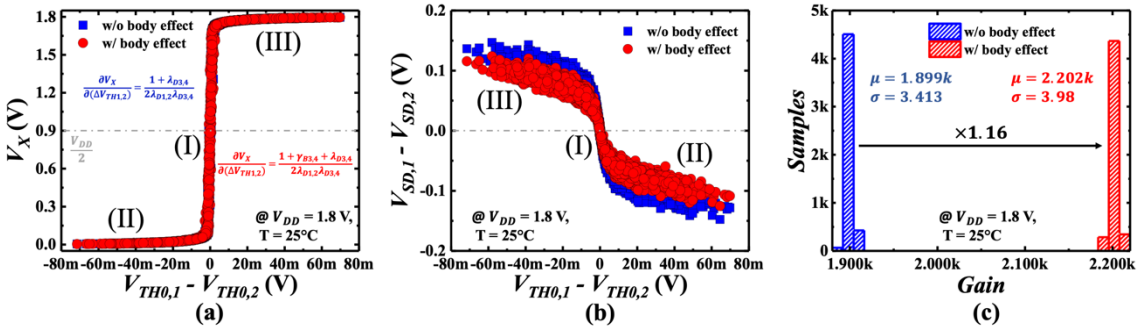


Fig. 3.48. Comparison at GK conditions (i.e., V_{DD} = 1.8 V and T = 25°C) between the two solutions in terms of (a)V_X voltage and (b) $V_{SD1} - V_{SD2}$ as function of the M1-M2 mismatch, and (c) gain distribution. Data came from 5k-run Monte Carlo simulations.

However, the effectiveness of this new design relies on reducing the M1-M2 strength difference sensitivity to the V_{DD} variations. Indeed, V_{SD1} and V_{SD2} are given by subtracting to V_{DD} and V_X the equations (3.21) and (3.22), respectively.

$$V_{SD1} = \frac{1}{n_{1,2} + n_{1,2}\lambda_{D3,4} + n_{3,4}\lambda_{D1,2} + n_{1,2}\gamma_{B3,4}}\left[(n_{1,2}\lambda_{D3,4})(V_{DD} - V_X) - n_{3,4}V_{TH0,1}(T) + n_{1,2}V_{TH0,3}(T) - n_{1,2}n_{3,4}V_T \ln\left(\frac{I_{0,1}W_1L_3}{I_{0,3}W_3L_1}\right)\right] \quad (3.25)$$

$$V_{SD2} = \frac{1}{n_{1,2} + n_{1,2}\lambda_{D3,4} + n_{3,4}\lambda_{D1,2} + n_{1,2}\gamma_{B3,4}}\left[(n_{1,2}\lambda_{D3,4})V_X - n_{3,4}V_{TH0,2}(T) + n_{1,2}V_{TH0,4}(T) - n_{1,2}n_{3,4}V_T \ln\left(\frac{I_{0,2}W_2L_4}{I_{0,4}W_4L_2}\right)\right] \quad (3.26)$$

From (3.25) and (3.26) the voltage drops across M1 and M2 of the circuit of Fig.3.47(c) show less voltage sensitivity than that of Fig. 3.47(b). However, as a said effect, the nominal voltage drops across them is lower. This means that using the same sizing for improving the area efficiency could lead to not enough M3 and M4 conductivities for always guarantying adequate voltage drops (i.e., above than 3−4 thermal voltages) across M1 and M2 thus resulting in a lower M1-M2 strength difference and hence lower V_X spread. This is due to the exponential term in the square bracket of (3.1) for which low V_{SD} variations result in high current variations. Increasing the M3 and M4 channel width (i.e., $W_{3,4}$) can counteracts the previous effect.

Extensive Monte Carlo (MC) simulations were performed to assess the stability of the two designed PUF bitcells, showed in Fig. 3.47(b) and (c), across process corners (TT, SS, FF, SF and FS), voltages (V_{DD} ranging from 1.8 V down to 0.4 V) and temperatures (ranging from 0 °C up to 100 °C). The stability analysis involved quantifying the percentage of unstable bits, which includes 'noisy' and 'flipped' bits, at a given PVT corner. Noisy bits correspond to V_X voltages falling in the neighborhood of $V_{DD}/2$, thus resulting into potentially unstable bits at the output of the subsequent inverter. More precisely, the bits that fall in the undefined input region of the inverter, defined as $V_{IH} - V_{IL} + 2V_T$. The term $2V_T$ was introduced to take more effectively into account the effect of on-chip noise at different operating conditions. On the other, flipped bits refer to the number of bits that permanently flip at different environmental conditions. Before showing the stability results it is important to understand the behavior of the area-efficient inverter at different process corners under voltage and temperature variations.

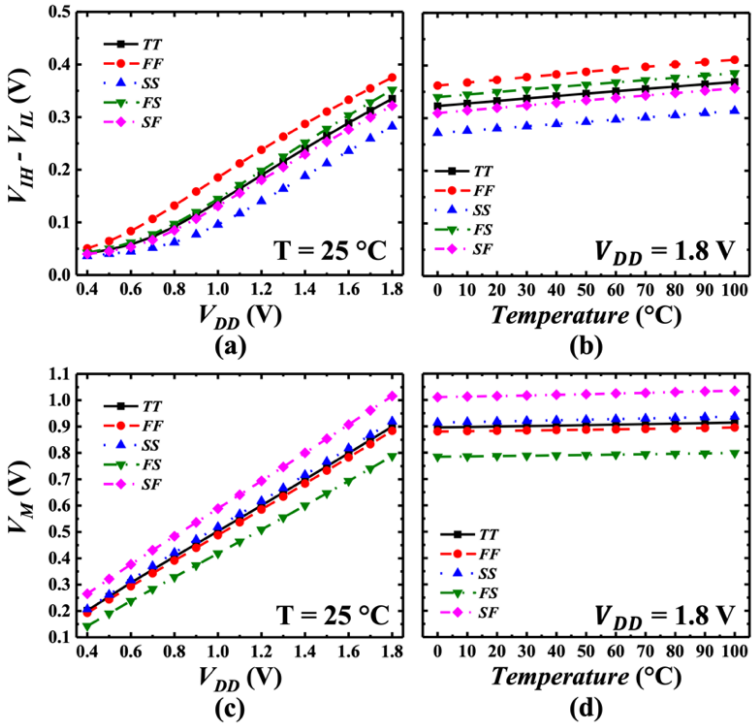


Fig. 3.49. Effect of voltage and temperature variations on the static parameters of the output inverter at different process corners. Effect of (a) voltage with T= 25 °C, and (b) temperature with V_{DD} = 1.8 V variations on the difference between minimum high- and maximum low-input voltages (V_{IH}-V_{IL}). Effect of (c) voltage with T= 25 °C and (d) temperature with V_{DD} = 1.8 V variations on the input logic-threshold (V_M).

Fig. 3.49(a)-(d) report the effect of process, voltage, and temperature variations on the amplitude of the unstable input region of the inverter as well as its input logic threshold. Fig. 3.49(a) and (b) illustrate difference between the minimum high- and maximum low-input voltages (i.e., V_{IH}-V_{IL}) as function of voltage (0.4−1.8 V) and temperature (0−100 °C) variations respectively. From Fig. 3.49(a), such a voltage difference exhibits a nearly linear decrease with decreasing V_{DD} down to 0.6 V, whereas the decreasing trend deviates from the linearity for V_{DD} below 0.6 V. In addition, Fig. 3.49(b) shows that V_{IH}-V_{IL} slightly increases with increasing the temperature. More precisely, in the corners TT, FS, and SF the static performances of the inverter in terms of V_{IH}-V_{IL} are quite close to each other with an increase of the amplitude at the corner FF (i.e., due to the higher transistor conductivities) and a decrease of the amplitude in the corner SS (i.e., due to lower transistor conductivities). From Fig. 3.49(c) the logic threshold increases linearly with V_{DD} keeping its value quite close to the mid-supply point at the corners TT, FF, SS with a slight increase and decrease at SF and FS corners respectively. From Fig. 3.49(d) the logic threshold exhibits a slight increase when increasing the temperature and shows the same trend of Fig. 3.49(c) under process variations. Fig. 3.50(a)-(d) show the percentage of unstable bits in the two circuits at different process corners, while separately considering the effect of voltage and temperature variations. bits averaged over process corners under voltage and temperature variations. Here, the unstable bits account for both ‘noisy’ and ‘flipped’ bits. The increase of the overall instability, compared to the bitcell described in the previous sub-section, is associated to the area-efficient inverter.

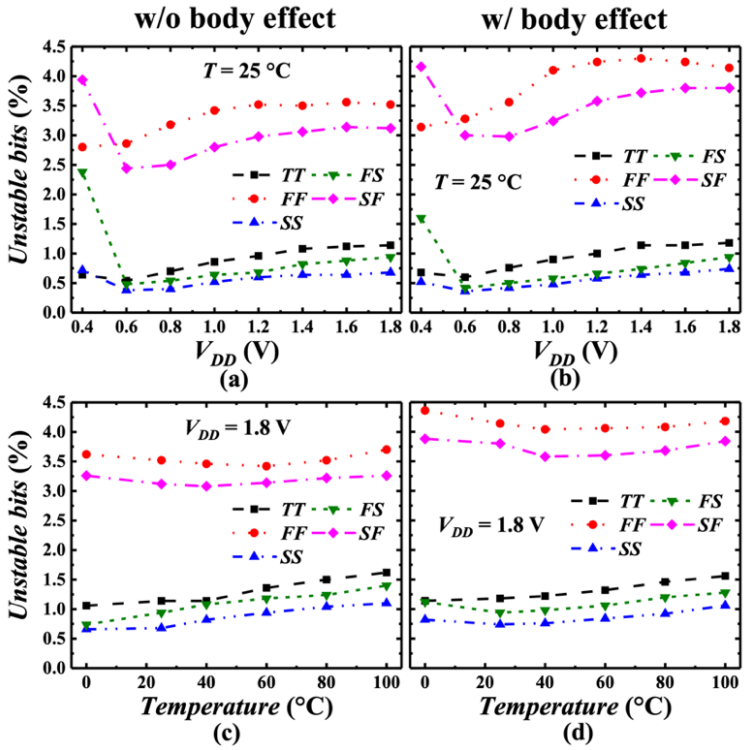


Fig. 3.50. Percentage of unstable bits ('noisy'+ 'flipped') at different process corners across voltages (T = 25 °C) and temperatures (V_{DD}= 1.8 V). (a) and (c) refer to the bitcell of Fig. 3.47(b), (b) and (d) refer to the bitcell of Fig. 3.47(c).

Anyway, both bitcell implementations exhibit higher bit instability at FF and SF corners. This can be ascribed to the fact that, referring to the core block, at these corners the increase in PMOS strength is more effective in minimum-sized M1-M2 than M3-M4, as shown in Fig. 3.28. This results in voltage drop across M1-M2 lower than $3-4V_T$, which counteracts the effect of their V_{TH} mismatch in differentiating the strength of two sub-circuits (i.e., lower $V_{SD1,2}$ variation is required for balancing the strength of the two sub-circuit) within the bitcell core, as stated above. The reduction of $V_{SD1,2}$ at FF and SF corners is emphasized in the design of Fig. 3.47(c) as given by the reverse body biasing that leads to increase the voltage drop on M3-M4, thus resulting into slightly higher percentage of unstable bits compared to its counterpart. This can be better appreciated in Fig. 3.51(a)-(b), which report the percentage of unstable. From these figures, the instability averaged across different process corners of the circuit in Fig. 3.47(b) is always lower than that of Fig. 3.47(c) (i.e., for the reasons stated above) even under voltage and temperature variations. Fig. 3.51(c) summarized the comparison in terms of unstable bits under different PVT conditions. From this figure, including the body effect within the core circuit results in a similar percentage of unstable bits than which does not include the body effect at the TT corner and GK conditions, i.e., 1.18% vs. 1.14%. Conversely, when averaging the unstable bits over process corners and different voltages/temperatures, the bitcell topology here proposed suffers from slightly higher bit instability than its counterpart. Indeed, when averaging across process corners at GK conditions the percentage of unstable bits increases from 1.88% up to 2.16%. On the other hand, when averaging across different process corners and voltages at T= 25 °C the instability increases from 1.73% up to 1.93%.

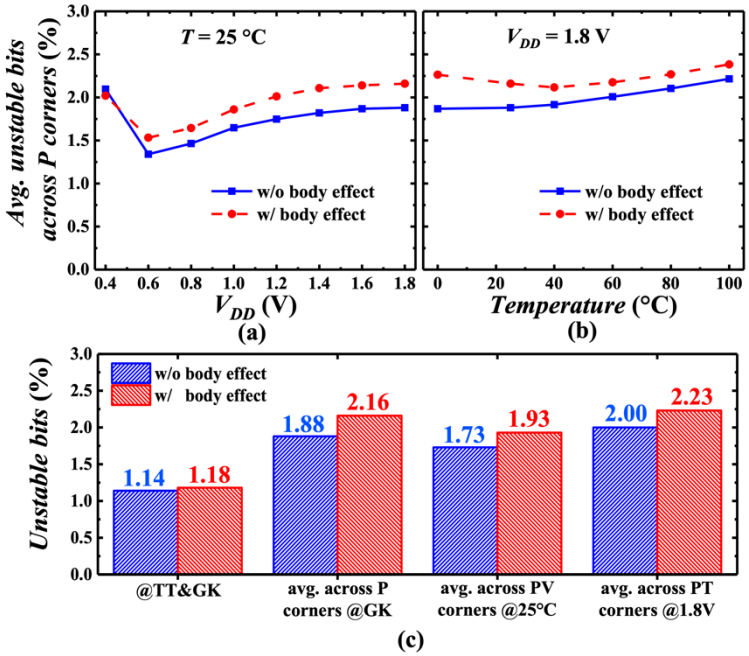


Fig. 3.51. Percentage of unstable bits averaged over process corners across (a) voltages and (b) temperatures for the two circuits of Fig. 3.47(b)-(c), (c) summary comparison in terms of unstable bits under different PVT conditions.

Finally, when averaging under temperature variations and different process corners at V_{DD}= 1.8 V the overall instability increases from 2.00% up to 2.23%. The comparison between the two circuits was extended by evaluating the BER and the static power consumption per bitcell across process corners, voltages and temperatures. Both figures of merit were estimated while considering an 8×32 bitcell array as implemented in the previous sub-section. The BER was evaluated accounting for 32-bit PUF output words. Fig. 3.52 summarizes the adopted procedure for estimating the BER as function of both the input−output characteristic of the inverter of Fig. 3.43(b) and the V_X distribution.

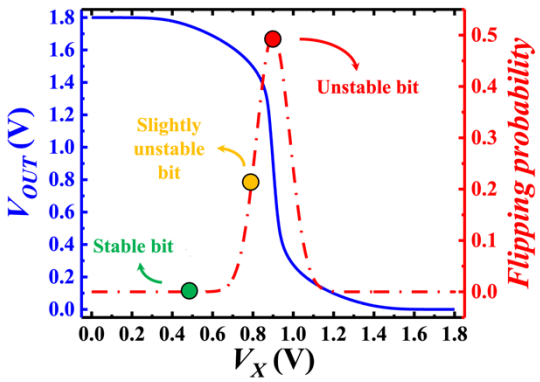


Fig. 3.52. Flipping probability as function of both input−output characteristic of the inverter of Fig.3.43(b) and V_X distribution.

From this figure, only the V_X samples that fall in the undefined input region of the inverter were considered as unstable with different probability depending on the proximity to the logic threshold of the inverter (i.e., V_M, which is the maximum-gain point). For example, if the V_X voltage is very close to the inverter logic threshold (the red dot in Fig. 3.52) its flipping probability is close to 0.5 which means that under different evaluations it could give both '0' and '1' with the same

probability. On the other hand, the orange dot represents a slightly unstable bit which means that under different evaluations in most cases that bit will give '1' as output and hence it could be easily filtered with a majority voting approach. Finally, the green dot represents a stable sample and hence noisy does not affect the output bit generation. In addition, the static power per bitcell was computed considering the effect of power gating within the PUF array, as enabled by the inverter design in Fig. 3.47(b)-(c). More specifically, the estimated power consumption takes into account that the output inverter is enabled only in one bitcell row at a time, whereas in the rest of bitcells its static power contribution is suppressed. Fig. 3.53 shows the comparison in terms of BER under different PVT conditions.

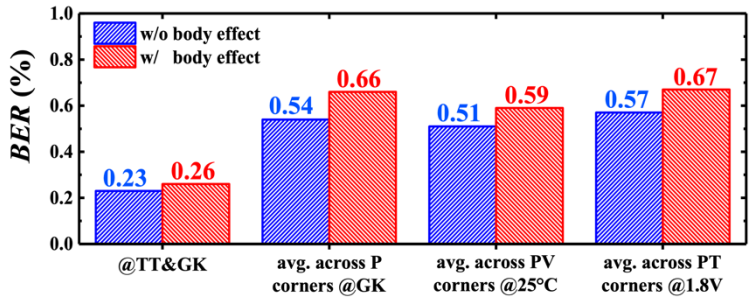


Fig. 3.53. Summary comparison between the two circuits of Fig. 3.47(b)-(c) in terms of bit error rate (BER) under different PVT conditions. Data refers to 32-bit PUF words within an 8×32 bitcell array.

From this figure, the observed trend is expectedly similar to that of the unstable bits in Fig. 3.51(c). Indeed, the two circuits reach similar BER at the TT corner and GK conditions, i.e., 0.23% and 0.26% respectively for the bitcell of Fig. 3.47(b) and (c). On the other hand, the new design, which includes the body effect, exhibits slightly higher BER than its counterpart when considering process variations, e.g., 0.66% vs. 0.54% when averaging over process corners at GK conditions. On the other hand, when averaging across different process corners and voltages at T= 25 °C the BER increases from 0.51% up to 0.59%. Finally, when averaging under temperature variations and different process corners at V_{DD}= 1.8 V the BER increases from 0.57% up to 0.67%. Fig. 3.54 shows the comparison in terms of static power per bitcell across PVT. It was computed considering the effect of power gating within the PUF array, as enabled by the inverter design in Fig. 3.47(b)-(c). More specifically, the estimated power consumption takes into account that the output inverter is enabled only in one bitcell row at a time, whereas in the rest of bitcells its static power contribution is suppressed. From this figure, we can observe slightly higher power consumption in the new design, which includes the body effect, compared to its counterpart. Indeed, the two circuits show similar power consumption at the TT corner and GK conditions, i.e., 98.4 nW and 103.5 nW, respectively for the bitcell of Fig. 3.47(b) and (c). On the other hand, the new design, exhibits slightly higher power consumption than that of Fig. 3.47(b) when considering process variations, e.g., 200.0 nW vs. 175.2 nW when averaging over process corners at GK conditions. On the other hand, when averaging across different process corners and voltages at T= 25 °C the power consumption increases from 51 nW up to 58.2 nW. Finally, when averaging under temperature variations and different process corners at V_{DD}= 1.8 V the power consumption increases from 183.3 nW up to 205.6 nW. This can be mainly ascribed to the higher percentage of unstable 'noisy' bits. Indeed, the latter correspond to the cases of intermediate V_X voltages generated by the bitcell core, which result into a noticeable increase in the dominant power contribution associated with the output inverter.

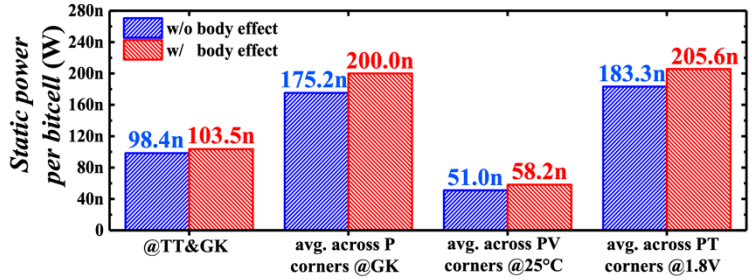


Fig. 3.54. Summary comparison between the two circuits of Fig. 3.47(b)-(c) in terms of static power consumption per bitcell under different PVT conditions. BER data refers to an 8×32 bitcell array with power gating.

Moreover, it is important to point out how the power gating helps for obtaining around the same power consumption of that reported in the previous sub-section thus counteracting the increase of the supply current due to the area-efficient inverter.

3.4.3 Comparison with prior works

Finally, Table III summarizes the results of the two simulated PUF circuits. The table also reports measurement results of relevant prior art CMOS PUFs, including the 4T voltage divider-based implementation originally proposed in [13]. When compared to the latter, the two circuits analyzed in this work shows ~2× higher percentage of unstable bits and BER. However, such better stability of the solution in [13] is achieved at the cost of significantly larger bitcell area (~7,200F^2), i.e., 2.1× and 3.4× higher than the design of Fig. 3.47(b) and (c), respectively. This because the output inverter in [13] was designed using longer channel devices (i.e., L = 2.5 µm) to ensure high gain and high robustness against process variations, as well as to keep its power consumption low in absence of power gating. When compared to other prior art solutions reported in Table III, the bitcell area of the circuit of Fig. 3.47(a) (Fig. 3.47(c)) is 3.9-6.9× (2.4-4.3×) higher than [30], [46]-[37], similar (1.7× lower) than [43], and 2.7× (4.4×) lower than [46].

TABLE III. SUMMARY RESULTS AND COMPARISON WITH STATE-OF-THE-ART CMOS PUF DESIGNS (DATA W/O STABILITY-ENHANCEMENT TECHNIQUES)

	Simulations		Measurements						
	This work original design Fig. 3.47(b)	This work new design Fig. 3.47(c)	JSSC 2022 [13]	JSSC 2021 [30]	JSSC 2017 [43]	JSSC 2018 [50]	ISSCC 2017 [46]	JSSC 2020 [47]	ISSCC 2018 [37]
tech. [nm]	180	180	180	130	14	40	180	65	180
bitcell type	4T voltage divider	4T voltage divider	4T voltage divider	SRAM	metastable	cascode current mirror	2T amplifier (DNW)	3T amplifier	leakage based (footed)
bitcell area [F^2]	3,440	2,140	7,222	497	9,388	3,644	553	562	890
native unstable bits [%]	1.14[a]	1.18[a]	0.61	2.71	26.8	2.55	1.67	2.95	5.62
unstable bits across process corners [%]	1.88[b]	2.16[b]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
native BER [%]	0.23[a]	0.26[a]	0.13	0.29	5.76	0.81	0.13	0.3	0.69
BER across process corners [%]	0.54[b]	0.66[b]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
static power per bitcell [W]	98.4n[a,c] (1.8V, 25°C)	103.5n[a,c] (1.8V, 25°C)	85.1n[a] (1.8V, 25°C)	N/A	3u (0.65V, 70°C)	8n[c] (0.9V, 25°C)	26p (1.2V, 25°C)	N/A	N/A

[a] @TT corner and GK conditions [b] @GK conditions [c] with power gating

Moreover, the performance of the two circuits in terms of native unstable bits and BER as estimated by simulations is quite competitive with respect to other designs. For instance, the percentage of unstable bits achieved by the circuit of Fig. 3.47(b) (Fig. 3.47(c)) is 1.5-23.5× (1.4-22.7×) lower than prior art, with the exception of [13] as discussed above. Moreover, the estimated static power per bitcell of the two analyzed circuits is in the same order of magnitude compared

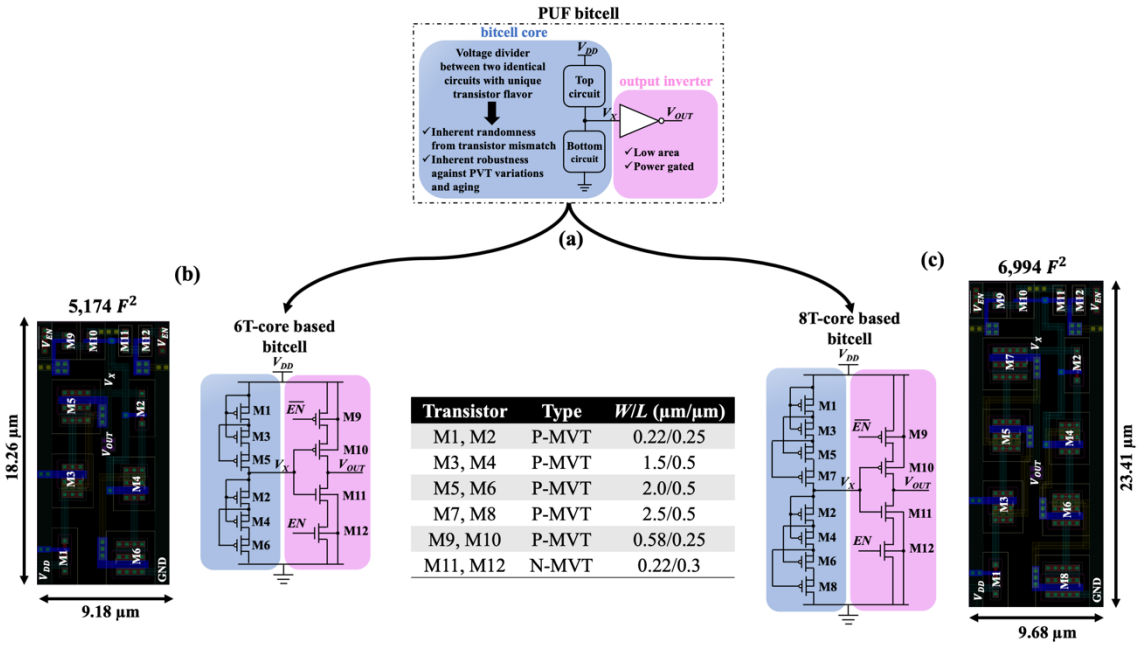
to that in [50] with power gating, ~30× lower than [43], and higher than [46] by more than one order of magnitude.

3.5 More stacked solutions

With the aim of further improving the overall stability, different circuital versions have been considered. In particular, comparison between simulations and measurements partially confirmed the effectiveness of reducing the percentage of samples that fall in the unstable region of the output inverter for improving the resiliency to the noise conditions. Despite the high gain provided by the 4T-core circuit, a small percentage (e.g., around 1%) of V_X samples still fall in the unstable region and hence result instable at the output of the inverter, due to on-chip noise. In this chapter we considered more complex sub-circuit architectures for increasing the overall gain on M1-M2 mismatch. A general message from the 4T-core analysis is that when adding one more stacked transistor in each sub-circuit the gain provided by the circuit increase of a factor $1/\lambda_D$ which allows making the V_X voltages well-readable, especially when the M1-M2 mismatch is small. This also reduces the current absorbed by the output inverter thus resulting in both stability and energy improvements. However, the main drawback is that each added transistor results in a larger increase of the bitcell footprint due to the layout constraints and the need of proper sizing the additional transistors for guarantying a correct operation.

3.5.1 Design guidelines of 6T-core and 8T-core

Fig. 3.55(a)-(c) illustrate the PUF design concept along with the schematic and layout of the proposed 6T-core based and 8T-core based solutions and the transistor sizing.



Transistor	Type	W/L (µm/µm)
M1, M2	P-MVT	0.22/0.25
M3, M4	P-MVT	1.5/0.5
M5, M6	P-MVT	2.0/0.5
M7, M8	P-MVT	2.5/0.5
M9, M10	P-MVT	0.58/0.25
M11, M12	N-MVT	0.22/0.3

Fig. 3.55. (a) Bitcell design concept along with schematic and layout (b) 6T-core based bitcell and (c) 8T-core based bitcell.

Compared to the 4T-core based solution, 6T and 8T cores include one and two, respectively, negative-V_{SG} transistors in each sub-circuit for further amplify the M1-M2 mismatch. Following the same guidelines described for the 4T-core, the added transistors must be sized larger for ensuring an adequate voltage drops on each transistor. Indeed, as shown in Fig. 3.55, M5-M8 were sized with the same channel length of M3 and M4 but with larger channel width. More precisely, M5 and M6 were sized with L= 0.5 µm and W= 2.0 µm while M7 and M8 with L= 0.5 µm and W = 2.5 µm. From an analytical point of view, following the same assumptions and procedures made for the 4T-core circuit, the V_X equations for the 6T-core and 8T-core solutions are given by

$$V_{X,6T} = \frac{V_{DD}}{2} + \frac{1+\lambda_{D5,6}(1+\lambda_{D3,4})}{2\lambda_{D1,2}\lambda_{D3,4}\lambda_{D5,6}}\left[V_{TH0,1} - V_{TH0,2} + k_{T1} - k_{T2} + n_{1,2}V_T \ln\left(\frac{I_{0,1}W_1L_2}{I_{0,2}W_2L_1}\right)\right] + \frac{1+\lambda_{D5,6}}{2\lambda_{D3,4}\lambda_{D5,6}}\left[V_{TH0,3} - V_{TH0,4} + n_{3,4}\ln\left(\frac{I_{0,3}W_3L_4}{I_{0,4}W_4L_3}\right)\right] + \frac{1}{2\lambda_{D5,6}}\left[V_{TH0,5} - V_{TH0,6} + \ln\left(\frac{I_{0,5}W_5L_6}{I_{0,6}W_6L_5}\right)\right] \quad (3.27)$$

$$V_{X,8T} = \frac{V_{DD}}{2} + \frac{1+\lambda_{D7,8}[1+\lambda_{D5,6}(1+\lambda_{D3,4})]}{2\lambda_{D1,2}\lambda_{D3,4}\lambda_{D5,6}\lambda_{D7,8}}\left[V_{TH0,1} - V_{TH0,2} + k_{T1} - k_{T2} + n_{1,2}V_T \ln\left(\frac{I_{0,1}W_1L_2}{I_{0,2}W_2L_1}\right)\right] + \frac{1+\lambda_{D7,8}(1+\lambda_{D5,6})}{2\lambda_{D3,4}\lambda_{D5,6}\lambda_{D7,8}}\left[V_{TH0,3} - V_{TH0,4} + n_{3,4}\ln\left(\frac{I_{0,3}W_3L_4}{I_{0,4}W_4L_3}\right)\right] + \frac{1+\lambda_{D7,8}}{2\lambda_{D5,6}\lambda_{D7,8}}\left[V_{TH0,5} - V_{TH0,6} + \ln\left(\frac{I_{0,5}W_5L_6}{I_{0,6}W_6L_5}\right)\right] + \frac{1}{2\lambda_{D7,8}}\left[V_{TH0,7} - V_{TH0,8} + \ln\left(\frac{I_{0,7}W_7L_8}{I_{0,8}W_8L_7}\right)\right] \quad (3.28)$$

Where, considering the adopted sizing the dominant variability is that of M1 and M2. This allows us neglecting the M3-M8 mismatch and hence writing (3.27) and (3.28) as follow

$$V_{X,6T} \approx \frac{V_{DD}}{2} + \frac{1+\lambda_{D5,6}(1+\lambda_{D3,4})}{2\lambda_{D1,2}\lambda_{D3,4}\lambda_{D5,6}}\left[V_{TH0,1} - V_{TH0,2} + k_{T1} - k_{T2} + n_{1,2}V_T \ln\left(\frac{I_{0,1}W_1L_2}{I_{0,2}W_2L_1}\right)\right] \quad (3.29)$$

$$V_{X,8T} \approx \frac{V_{DD}}{2} + \frac{1+\lambda_{D7,8}[1+\lambda_{D5,6}(1+\lambda_{D3,4})]}{2\lambda_{D1,2}\lambda_{D3,4}\lambda_{D5,6}\lambda_{D7,8}}\left[V_{TH0,1} - V_{TH0,2} + k_{T1} - k_{T2} + n_{1,2}V_T \ln\left(\frac{I_{0,1}W_1L_2}{I_{0,2}W_2L_1}\right)\right] \quad (3.30)$$

where in absence of mismatch (e.g., $V_{TH0,1} - V_{TH0,2} = 0$) the V_X voltages are equal to the mid-supply point. When mismatch occurs, the top and bottom circuits push the V_X voltage as far from $V_{DD}/2$ as low is the equivalent DIBL of the stack which is much lower compared to that provided by the 4T-core circuit. As result both 6T and 8T circuits provide well readable V_X voltages even with small mismatch. Fig. 3.56(a) and (b) report the scatter plot between the V_X voltage as function of the M1-M2 mismatch for both 6T-core and 8T-core circuits and the distribution of the gain of the circuits. Data came from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T= 25°C). Fig. 3.56(a) illustrates the V_X sensitivity of both circuits to the M1-M2 mismatch, from which it is notable the ability of the circuit of generating always well readable V_X voltages even in small mismatch conditions.

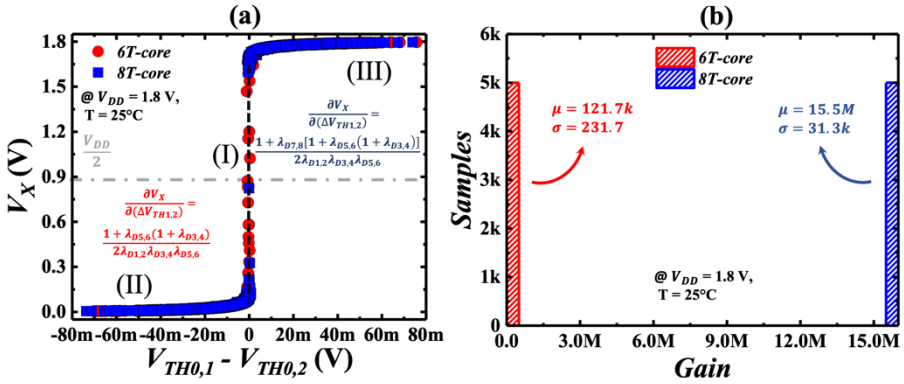


Fig. 3.56. Comparison at GK conditions (i.e., V_{DD} = 1.8 V and T = 25°C) between 6T-core and 8T-core circuits in terms of (a)V_X voltage as function of the M1-M2 mismatch, and (b) gain distribution. Data came from 5k-run Monte Carlo simulations.

This is confirmed by Fig. 3.56(b) which shows the statistical distributions of the gain of circuits of Fig. 3.55(b) and (c). From this figure the mean gain value is of 121.7k and 15.5M for 6T-core and 8T-core circuit, respectively. For better appreciating the circuit behavior, Fig. 3.57(a)-(d) provide the voltage repartition in the 6T and 8T voltage dividers in the cases of small mismatch. More precisely, Fig. 3.57(a) and (b) illustrate the voltage drops across each transistor in the voltage divider in the case of weak logic '0' (e.g., $|V_{TH0,1}|$ slightly higher than $|V_{TH0,2}|$) for the 6T-core and 8T-core circuits, respectively, whereas Fig. 3.57(c) and (d) illustrate the voltage drops in the case of weak logic '1' (e.g., $|V_{TH0,2}|$ slightly higher than $|V_{TH0,1}|$).

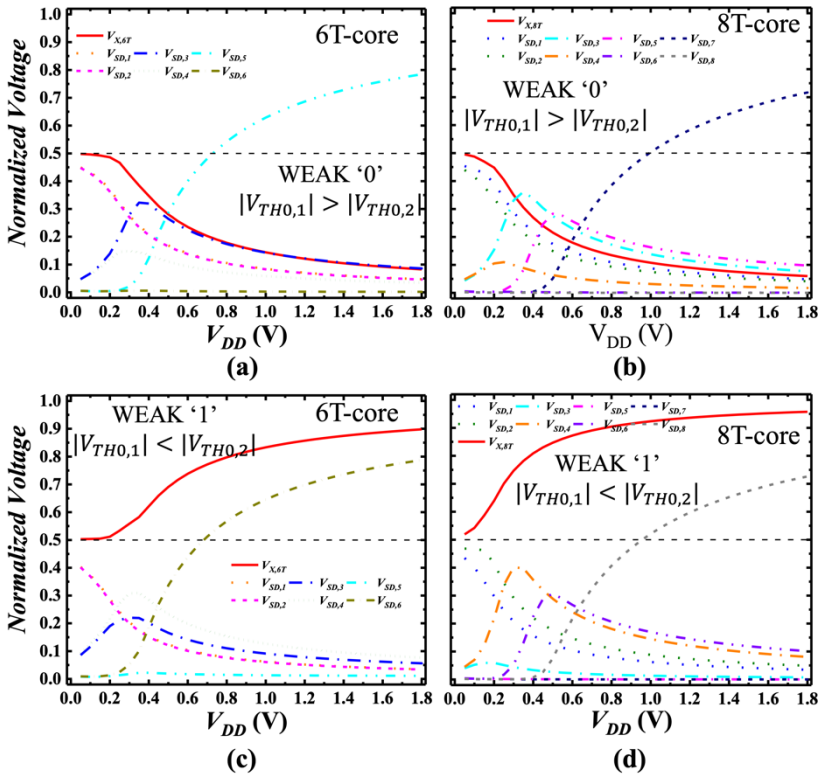


Fig. 3.57. Voltage drops across the transistors in the (a) and (c) 6T-core voltage divider and (b) and (d) 8T-core voltage divider, normalized to the V_{DD} for weak M1-M2 mismatch. (a) and (b) weak logic '0', (c) and (d) weak logic '1'.

Fig. 3.57 carries out two important features: (*i*) the transistors in the stack effectively translate small mismatch in large V_X deviation from the mid-supply point; and (*ii*) the more stacked transistors (e.g., M5-M8) do not affect the correct operation at very low voltages. Indeed, if proper sized, when decreasing the V_{DD} below 0.6 V their voltage drops gradually approach to zero. This can be ascribed to their large conductivity compared to the other in the stack. For example, as shown in Fig. 3.57(d) in the 8T-core circuit for V_{DD} values below 0.5 V, the voltage drops across M5 and M6 is not large enough for reducing the M7 and M8 conductivity enough to require large V_{SD} for delivering the same current. This means that these circuits can always guarantee a correct operation under voltage variation regardless of the number of transistors in each sub-circuit. Indeed, at very low voltage the more stacked transistors act as short circuits due to their higher conductivity. This allows improving the overall performance at higher V_{DD} compared to the 4T-core, while preserving its resiliency at very low voltage operations.

3.5.2 Simulations and measurements of 6T-core and 8T-core

As result, Fig. 3.58 (a)-(d) provide both the measured V_X and I_{core} trends as function of V_{DD} variations for both 6T-core and 8T-core circuits across 20 dice.

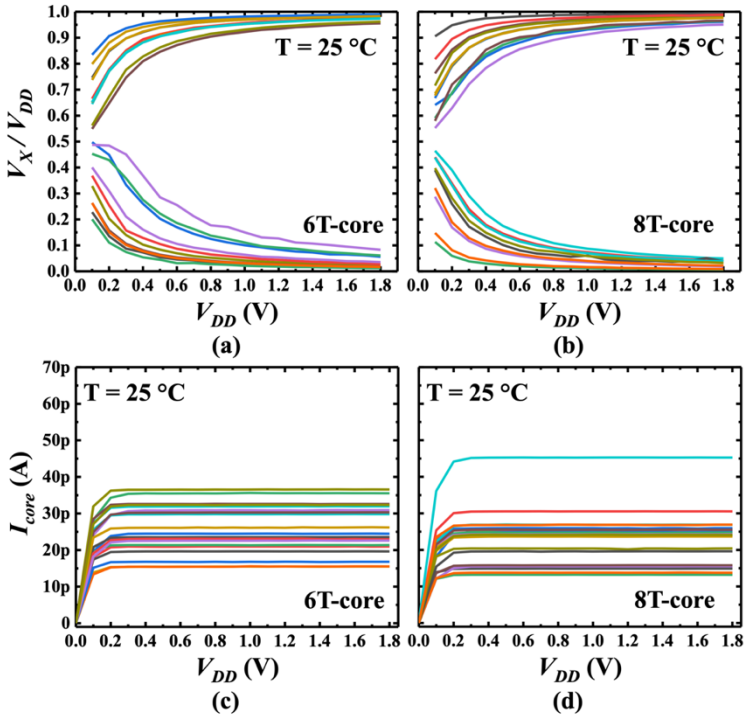


Fig. 3.58. Measured V_X voltage normalized to V_{DD} of (a) 6T-core (i.e., $V_{X,6T}$) and (b) 8T-core (i.e., $V_{X,8T}$) circuits under voltage variations at T= 25 °C across 20 samples. Absorbed current from (c) 6T-core (i.e., I_{6T}) and (d) 8T-core (i.e., I_{8T}) circuits under voltage variations at T= 25 °C across 20 samples.

Fig.3.58 (a) and (b) provide the normalized V_X trend as function of the voltage variations for 6T-core and 8T-core circuits, respectively. These figures highlight the ability of the proposed solutions of pushing the V_X voltages far from the mid-supply point, regardless of the considered voltage, even in small mismatch conditions. From Fig. 3.57, when considering a supply voltage of 100 mV the negative-V_{SG} transistors act as short circuits, thus indicating that the output voltage

is only function of M1 and M2. In this regard, such voltage (i.e., $V_{X,6T}$) shows a deviation from $V_{DD}/2$ which is proportional to the M1-M2 mismatch (i.e., $V_X - V_{DD}/2$ is proportional to $\Delta V_{TH0}/2\lambda_{D1,2}$). This means that at very low supply voltage the V_X voltages which are quite close to the mid-supply point indicate the circuit is operating in condition of small mismatch (i.e., $V_{TH0,1} \approx V_{TH0,2}$). As result, Fig. 3.58(a) shows two V_X samples which are quite close to the mid supply point at V_{DD}= 0.1 V, thus potentially unstable bits under voltage and temperature variations. However, the same figure proves the effectiveness of the proposed solution of shielding the M1-M2 mismatch from the voltage variations. Fig. 3.58(c) and (d) provide the supply current of both 6T-core and 8T-core circuits as function of the voltage variations. From these figures, above a certain voltage the absorbed current becomes insensitive to the voltage variations, thus proving the effectiveness of the proposed solutions in reducing the equivalent DIBL of the stack. Indeed, from Fig. 3.58(c) and (d), both the averaged 6T-core and 8T-core supply currents (dissipated power) decrease from 25.6 pA (46.08 pW) to 25.57 pA (10.23 pW) and from 23.288 pA (i.e., 41.92 pW) to 23.285 pA (i.e., 9.32 pW), respectively, when reducing the V_{DD} from 1.8 V down to 0.4 V. On the other hand, Fig. 3.59(a)-(d) provide both the measured V_X and I_{core} trends as function of temperature variations for both 6T-core and 8T-core circuits across 20 dice.

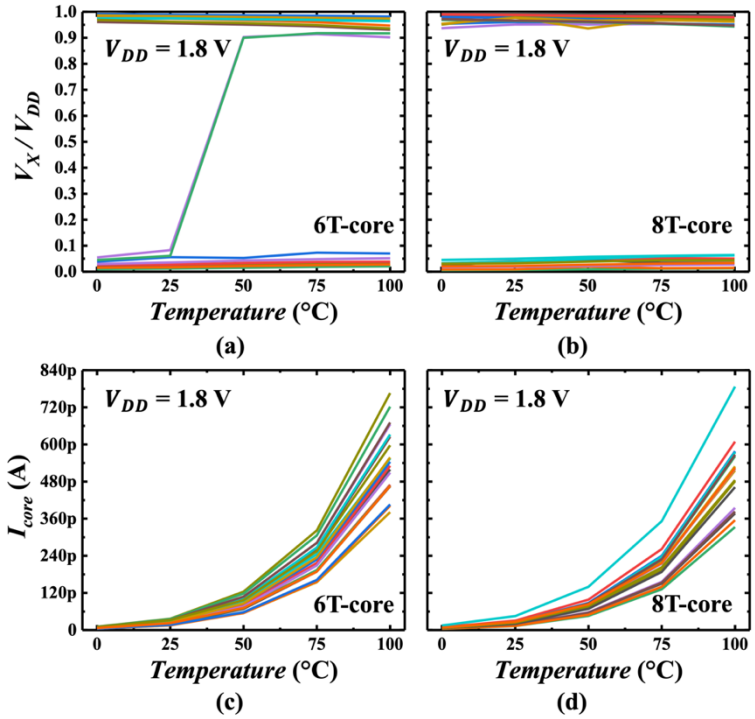


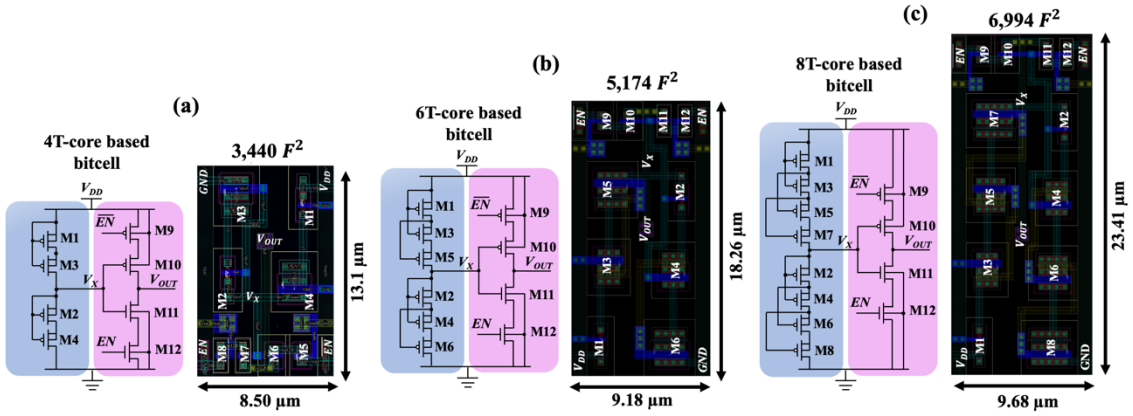
Fig. 3.59. Measured V_X voltage normalized to V_{DD} of (a) 6T-core (i.e., $V_{X,6T}$) and (b) 8T-core (i.e., $V_{X,8T}$) circuits under temperature variations at V_{DD}= 1.8 V across 20 samples. Absorbed current from (c) 6T-core (i.e., I_{6T}) and (d) 8T-core (i.e., I_{8T}) circuits under temperature variations at V_{DD}= 1.8 V across 20 samples.

Fig. 3.59(a) and (b) provide the measured V_X voltage normalized to V_{DD} for both 6T-core and 8T-core circuits under temperature variations. Fig.3.59(a) shows that two bits flipped under temperature variations. Indeed, they correspond to the samples whose V_X voltage was very close to the mid-supply point at V_{DD}= 0.1 V mentioned in the above analysis. This proves that these solutions can counteract the effect of the mismatch in terms of the DIBL coefficients (i.e., $\lambda_{D,1}$

and $\lambda_{D,2}$), but the output voltage can still exhibit an instable behavior under temperature variations due to the mismatch in terms of the temperature coefficient (i.e., $k_{T,1}$ and $k_{T,2}$). Anyway, all the V_X voltages which do not flip remain quite close to the two edges (i.e., V_{DD} and ground) under temperature variations thus potentially reducing the percentage of noisy bits during the readout phase. On the other hand, Fig. 3.59(c) and (d) report the measured 6T-core and 8T-core supply currents under temperature variations. From these figures, both currents exhibit an exponential relationship with the temperature variations due to the deep sub-threshold operations. Indeed, from Fig. 3.59(c) and (d), both the averaged 6T-core and 8T-core supply currents (dissipated power) increase from 25.6 pA (46.08 pW) to 547.7 pA (985.9 pW) and from 23.288 pA (i.e., 41.92 pW) to 506.8 pA (i.e., 912.24 pW), respectively, when increasing the temperature from 25°C up to 100°C.

3.5.3 Simulation results of the 6T-core and 8T-core based bitcells

These two solutions along with the 4T-core based bitcell were simulated through extensive Monte Carlo simulations across different PVT corners. Where, for safe comparison, the considered 4T-core based bitcell is that of Fig. 3.47(a). Fig. 3.60(a)-(c) illustrate the 4T, 6T, and 8T bitcells, respectively, along with the adopted sizing and the respective layout.



Transistor	Type	*W/L* (µm/µm)	Transistor	Type	*W/L* (µm/µm)	Transistor	Type	*W/L* (µm/µm)
M1, M2	P-MVT	0.22/0.25	M5, M6	P-MVT	2.0/0.5	M9, M10	P-MVT	0.58/0.25
M3, M4	P-MVT	1.5/0.5	M7, M8	P-MVT	2.5/0.5	M11, M12	N-MVT	0.22/0.3

Fig. 3.60. Schematic and Layout of the (a) 4T-core based, (b) 6T-core based, and (c) 8T-core based PUF bitcell.

This figure carries out that the main drawback of increasing the stack of each sub-circuit within the bitcell core is the increase of the bitcell footprint. Indeed, compared to the 4T-core based solution, the 6T-core and 8T-core based bitcells degrade the occupied area of 50.41% and 103.31%, respectively. Fig. 3.61(a)-(c) provide the statistical distribution of the V_X voltage of the 4T-core, 6T-core, and 8T-core solutions, respectively from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C) at TT corner. From this figure is notable how passing from the 4T-core to the 6T- and 8T-core solutions allows decreasing the percentage of potentially noisy bits (i.e., the percentage of bits that fall in the unstable input region of the subsequent inverter) at GK conditions from 1.14% down to 0.08% and 0%, respectively.

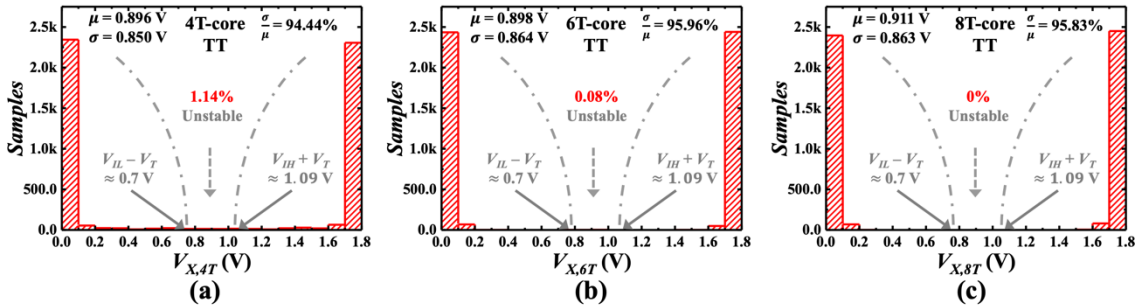


Fig. 3.61. Statistical distribution of the V_X voltage of the (a) 4T-cor, (b) 6T-core, and (c) 8T-core from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T = 25°C) at TT corner.

Fig. 3.62 (a)-(f) provide simulated trends of the unstable bits for the 4T-core, 6T-core, and 8T-core based solutions, respectively under both voltage and temperature variations at TT corner. This figure also illustrates the trend of both noisy and flipped bits under voltage, Fig. 3.62 (a)-(c), and temperature, Fig. 3.62(d)-(f), variations. Noisy bits correspond to V_X voltages falling in the neighborhood of $V_{DD}/2$, thus resulting into potentially unstable bits at the output of the subsequent inverter. More precisely, the bits that fall in the undefined input region of the inverter, defined as $V_{IH} - V_{IL} + 2V_T$. The term $2V_T$ was introduced to take more effectively into account the effect of on-chip noise at different operating conditions. On the other, flipped bits refer to the number of bits that permanently flip at different environmental conditions.

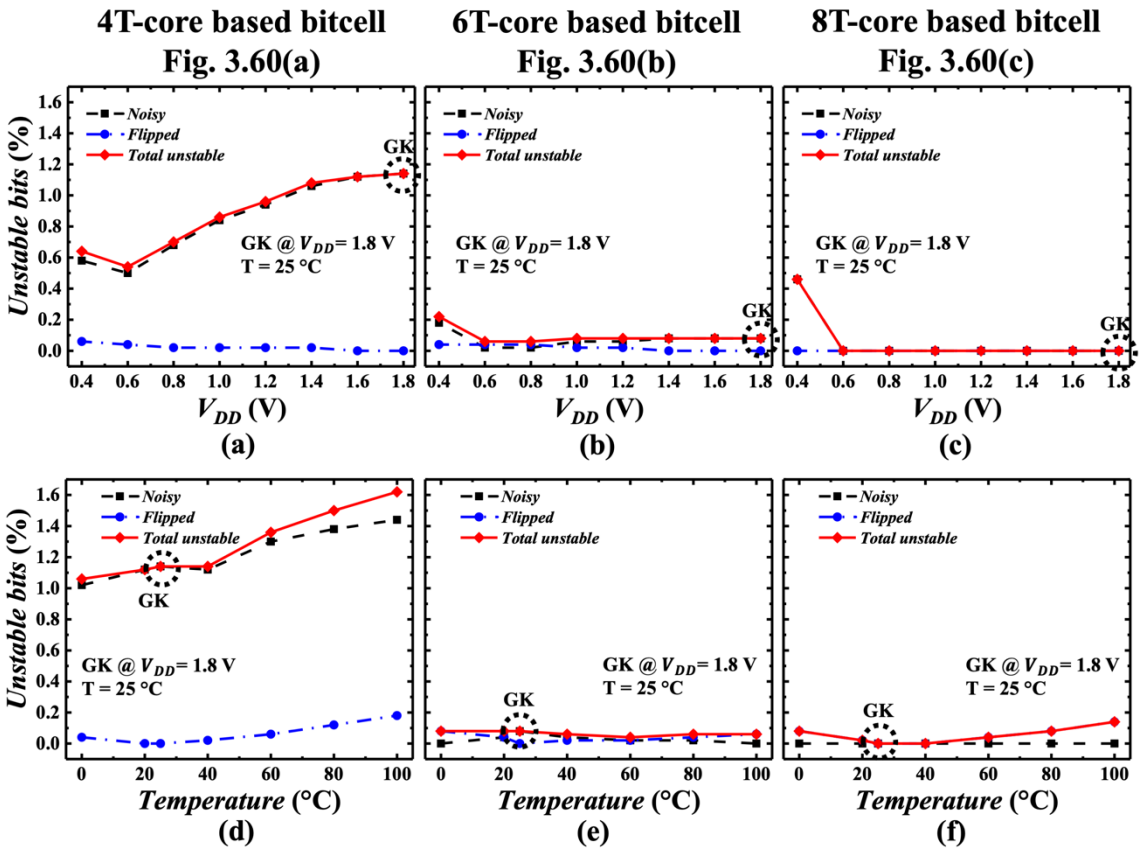


Fig. 3.62. Percentage of simulated unstable bits for the (a) and (d) 4T-core based, (b) and (e) 6T-core based, and (c) and (f) 8T-core based solutions from 5k-run Monte Carlo simulations at TT corner under (a)-(c) V_{DD} variations at T= 25 °C and (d)-(f) temperature variations at V_{DD}= 1.8 V.

Fig. 3.62 (a) and (d) report the stability trend of the 4T-core based bitcell of Fig. 3.60(a) under both voltage and temperature variations, respectively. From Fig. 3.62(a), the overall instability under voltage variations is dominated by noisy bits. Indeed, the percentage of noisy bits decreases from 1.14% down to 0.58% when decreasing the V_{DD} from 1.8 V down to 0.4 V whereas the percentage of flipped bits increases up to 0.06% under the same V_{DD} variations. From Fig. 3.62(d) the instability under temperature variations is mainly due to the increase of noisy bits. Indeed, the percentage of noisy bits increases from 1.14% up to 1.44% when increasing the temperature from 25 °C up to 100 °C whereas the percentage of flipped bits increases up to 0.18% under the same temperature variations. Fig. 3.62(b) and (e) provide the stability trend of the 6T-core based bitcell of Fig. 3.60(b) under both voltage and temperature variations. From Fig. 3.62(b) the stability under voltage variations is improved, compared to the 4T-core solutions. Indeed, the percentage of noisy bits increases from 0.06% up to 0.18% when decreasing the V_{DD} down to 0.4 V, whereas the percentage of flipped bits increases up to 0.04% under the same V_{DD} variations. From Fig. 3.62(e) the instability under temperature variations is mainly due to the increase of flipped bits. Indeed, the percentage of noisy bits remains quite constant under temperature variations, whereas the flipped bits increase up to 0.06% when increasing the temperature up to 100 °C. Finally, Fig. 3.62(c) and (f) report the trend of the unstable bits of the 8T-core based bitcell of Fig. 3.60(c) as function of both voltage and temperature variations, respectively. From Fig. 3.62(c) the percentage of unstable bits is 0% with V_{DD} ranging from 0.6 V to 1.8 V while increases up to 0.46% when decreasing the V_{DD} down to 0.4 V due to the increase of noisy bits. Fig. 3.62(f) provides the instability trend as function of the temperature variations. From this figure, the percentage of unstable bits increases up to 0.14% when increasing the temperature up to 100 °C due to the increase of flipped bits. Extensive Monte Carlo simulations were performed to assess the stability of the three designed PUF bitcells across process corners (TT, SS, FF, SF and FS), voltages (V_{DD} ranging from 1.8 V down to 0.4 V) and temperatures (ranging from 0 °C up to 100 °C). Fig. 3.63(a)-(f) provide the percentage of unstable bits (i.e., the sum of noisy and flipped bits) under voltage and temperature variations across different process corners for the three bitcells illustrated in Fig. 3.60(a)-(c). Fig. 3.63(a) and (d) report the trend of unstable bits under voltage and temperature variations, respectively, of the 4T-core based bitcell. From Fig. 3.63(a) the overall instability trend across different process corners is quite similar to that achieved at TT corner. However, the worst corners are FF and SF due to the higher pMOS conductivity. Indeed, at FF (SF) corner the percentage of unstable bits is 3.52% (3.12%) at V_{DD}= 1.8 V and 2.80% (3.94%) at V_{DD}= 0.4 V. From Fig. 3.63(d) the overall instability increases when varying the temperature from its value at GK conditions (e.g., 25 °C). Anyway, similar to the voltage variations, the worst corners are FF and SF corners. Indeed, at FF (SF) corner the percentage of unstable bits increases up to 3.7% (3.26%) at T= 100 °C. Fig. 3.63(b) and (e) report the trend of unstable bits under voltage and temperature variations, respectively, of the 6T-core based bitcell. From Fig. 3.63(b) the overall instability under voltage variations is much lower compared to that of Fig. 3.63(a). Like the 4T-core solution, the worst corners are FF and SF where the percentages of unstable bits are 0.42% and 0.24%, respectively, at V_{DD}= 1.8 V. On the other hand, at V_{DD}= 0.4 V the worst corners are SF and FS, due to the inverter behavior at these corners, where the percentages of unstable bits are 3.02% and 2.52%, respectively. From Fig. 3.63(e) the overall instability increases under temperature variations, especially in FF and SF corner, where the percentages of unstable bits are 1.06% and 0.76%, respectively, at T= 100 °C.

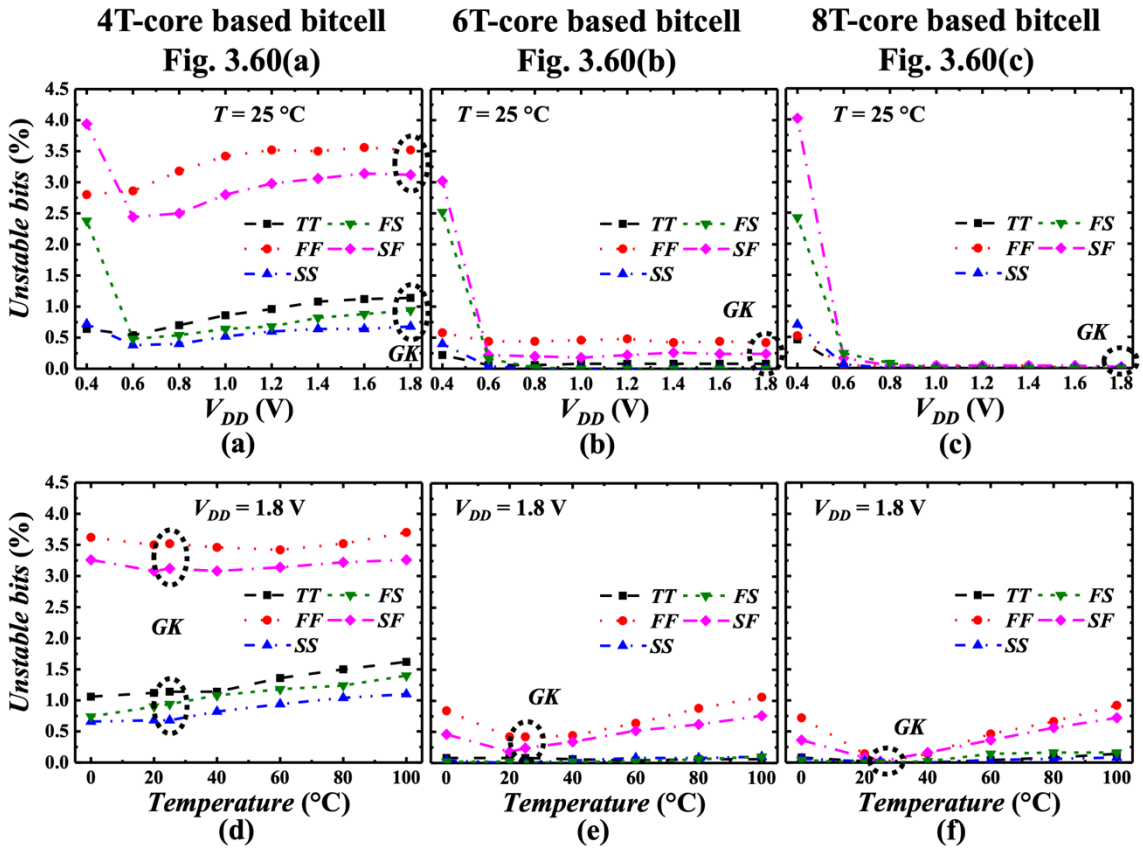


Fig. 3.63. Percentage of unstable bits ('noisy'+ 'flipped') at different process corners across (a)-(c) voltages (at T = 25 °C) and (d)-(f) temperatures (at V_{DD} = 1.8 V) from 5k-run Monte Carlo simulations. (a) and (d) refer to the bitcell of Fig. 3.60 (a), (b) and (e) refer to the bitcell of Fig. 3.60(b), (c) and (f) refer to the bitcell of Fig. 3.60(c).

Finally, Fig. 3.63(c) and (f) report the overall instability under voltage and temperature variations, respectively, of the 8T-core based bitcell. From Fig. 3.63(c) the high gain provided by the core circuit ensure a nearly-zero instability regardless of the process variations with V_{DD} ranging from 0.6 V to 1.8 V. The instability increases when V_{DD} approaches to 0.4 V, especially at SF and FS corners, where the percentages of unstable bits are 4.02% and 2.42%, respectively. From Fig. 3.63(f), the overall instability trend is quite similar to that of Fig. 3.63(e). Indeed, the instability increases when varying the temperature from the golden value (e.g., T= 25 °C), especially at FF and SF corners where the percentages of unstable bits are 0.92% and 0.72%, respectively, at T= 100 °C.

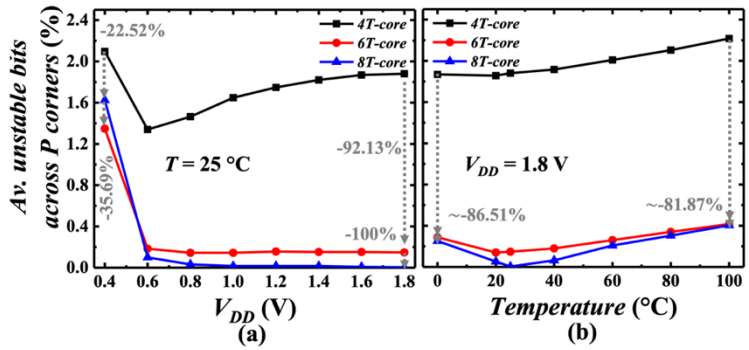


Fig. 3.64. Percentage of unstable bits averaged over process corners across (a) voltages and (b) temperatures for the three circuits of Fig. 3.60(a)-(c).

Furthermore, the benefits of the additional transistors on the overall instability at different PVT corners can be better appreciated in Fig. 3.64(a) and (b), which report the percentage of unstable bits averaged over process corners, under voltage and temperature variations. More precisely, Fig. 3.64(a) provides a summary comparison of the overall instability of the three solutions in Fig. 3.60(a)-(c). This figure highlights the benefits of adopting a more stacked solution, especially for V_{DD} ranging from 0.6 V to 1.8 V. Indeed, at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C) the percentage of unstable bits decreases from 1.88% down to 0.15% when moving from a 4T-core to a 6T-core based solution down to 0% when moving toward an 8T-based solution. However, for V_{DD} approaching to 0.4 V the instability showed by the three solutions of Fig. 3.60(a)-(c) is quite close to each other. This can be ascribed to the higher conductivity of the additional transistors which causes them acting as short circuits at very low supply voltages, as shown in Fig. 3.57(a)-(d). Indeed at V_{DD}= 0.4 V the percentage of unstable bits decrease from 2.1% down to 1.62% when moving from a 4T-core based to an 8T-core based solution and down to 1.35% when using a 6T-core based solution. Fig. 3.64(b) report the comparison of the overall instability of the three solutions in Fig. 3.60(a)-(c). From this figure, the instability trend of the three solutions is quite similar to each other. More precisely, 6T-core and 8T-core based solutions show a very close instability thanks to the suppression of noisy bits. Indeed, when moving from the 4T-core toward more stacked solutions the overall instability at 0 °C and 100 °C decreases from 1.87% down to ∼0.26% and from 2.22% down to ∼0.41%, respectively. For taking into account the effectiveness of the power gated inverter, the static power consumption was estimated considering an 8×32 bitcell array as implemented in the previous sub-sections. More specifically, the estimated power consumption takes into account that the output inverter is enabled only in one bitcell row at a time, whereas in the rest of bitcells its static power contribution is suppressed. Fig. 3.65(a)-(f) provide the average supply current (i.e., the sum of the current absorbed by the core circuit and that absorbed by the output inverter) at PVT corners for the three bitcells of Fig. 3.60(a)-(c). Data come from 5k-run Monte Carlo simulations. Fig. 3.65(a) and (d) illustrate the trend of the supply current under voltage and temperature variations, respectively, of the 4T-core based solution of Fig. 3.60(a). From Fig. 3.65(a) the supply current exhibits an exponential relationship with the supply voltage. Moreover, the worst corner is the FF, due to the lower V_X spread, where the supply current (dissipated power) at V_{DD}= 1.8 V and 0.4 V increases, respectively, from 54.7 nA (98.46 nW) up to 258 nA (464.4 nW) and from 229 pA (91.6 pW) up to 2.39 nA (956 pW). On the other hand, the best corner is the SS, due to the higher V_X spread, where the supply current (dissipated power) at V_{DD}= 1.8 V and 0.4 V decreases, respectively, from 54.7 nA (98.46 nW) down to 22.1 nA (39.78 nW) and from 229 pA (91.6 pW) down to 25.8 pA (10.32 pW). From Fig. 3.65(d) the supply current shows a nearly constant behavior under temperature variations. Like Fig. 3.65(a) the worst corner is the FF where the supply current (dissipated power) at T= 100 °C increases from 78.2 nA (140.76 nW) up to 265 nA (477 nW). On the other hand, the best corner is the SS where the supply current (dissipated power) at T= 100 °C decreases from 78.2 nA (140.76 nW) down to 33 nA (59.4 nW). Fig. 3.65(b) and (e) illustrate the trend of the supply current under voltage and temperature variations, respectively, of the 6T-core based solution of Fig. 3.60(b). From Fig. 3.65(a) the supply current exhibits a less pronounced exponential relationship with the supply voltage.

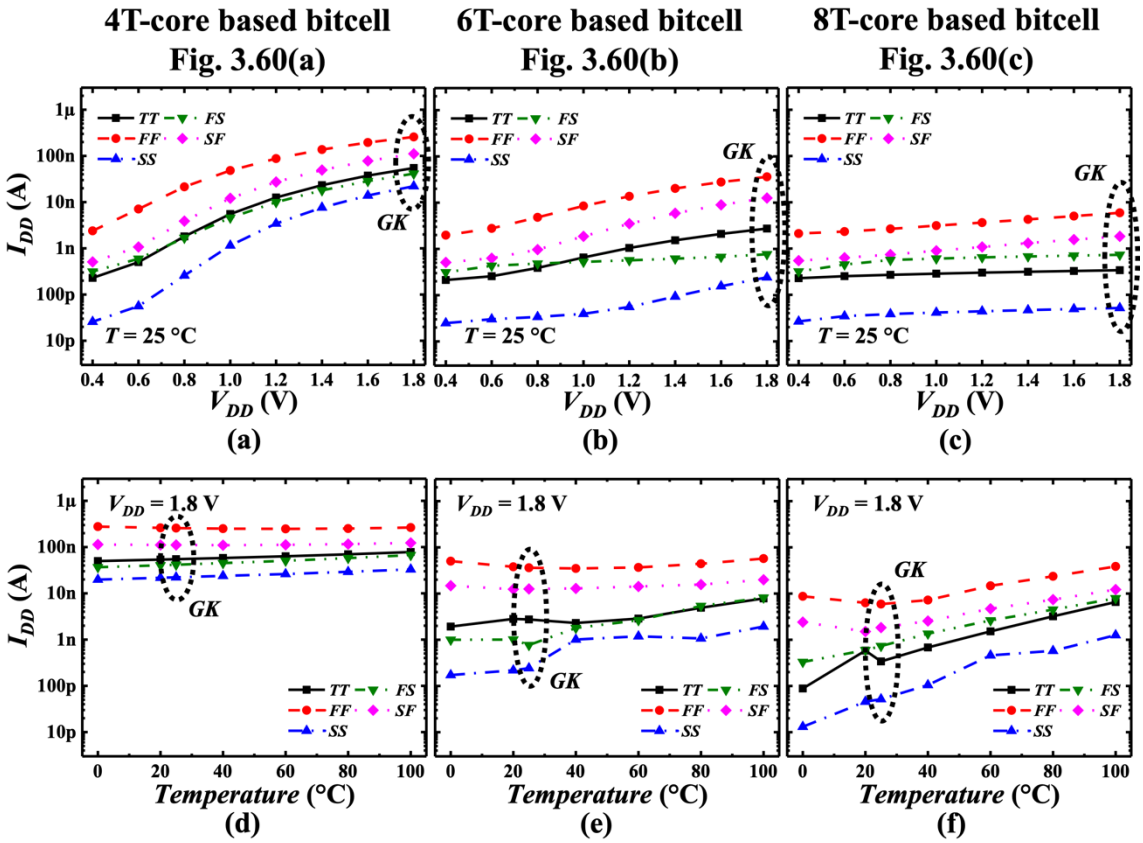


Fig. 3.65. Simulated supply current at different process corners across (a)-(c) voltages (at T = 25 °C) and (d)-(f) temperatures (at V_{DD}= 1.8 V) from 5k-run Monte Carlo simulations. (a) and (d) refer to the bitcell of Fig. 3.60 (a), (b) and (e) refer to the bitcell of Fig. 3.60(b), (c) and (f) refer to the bitcell of Fig. 3.60(c).

The behavior against process variations is quite similar to that showed by the 4T-core solution. Indeed, the worst corner is the FF, due to the lower V_X spread, where the supply current (dissipated power) at V_{DD}= 1.8 V and 0.4 V increases, respectively, from 2.75 nA (4.95 nW) up to 36.2 nA (65.16 nW) and from 213 pA (85.2 pW) up to 1.98 nA (792 pW). On the other hand, the best corner is the SS, due to the higher V_X spread, where the supply current (dissipated power) at V_{DD}= 1.8 V and 0.4 V decreases, respectively, from 54.7 nA (98.46 nW) down to 22.1 nA (39.78 nW) and from 213 pA (85.2 pW) down to 24.9 pA (9.96 pW). From Fig. 3.65(e) the supply current shows a nearly constant behavior under temperature variations. The worst corner is the FF where the supply current (dissipated power) at T= 100 °C increases from 7.85 nA (14.13 nW) up to 57.2 nA (102.96 nW). On the other hand, the best corner is the SS where the supply current (dissipated power) at T= 100 °C decreases from 7.85 nA (14.13 nW) down to 1.94 nA (3.49 nW). Finally, Fig. 3.65(c) and (f) report the trend of the supply current under both voltage and temperature variations, respectively, of the 8T-core solution of Fig. 3.60(c). From Fig. 3.65(c) the supply current exhibits a nearly constant trend under voltage variations. Moreover, the worst corner is the FF where the supply current (dissipated power) at V_{DD}= 1.8 V and 0.4 V increases, respectively, from 340 pA (612 pW) up to 5.92 nA (10.66 nW) and from 228 pA (91.2 pW) up to 2.11 nA (844 pW). On the other hand, the best corner is the SS, due to the higher V_X spread, where the supply current (dissipated power) at V_{DD}= 1.8 V and 0.4 V decreases, respectively, from 340 pA (612 pW) down to 51.9 pA (93.42 pW) and from 228 pA (91.2 pW) down to 26.6 pA (10.64 pW). From Fig. 3.65(f) the supply current shows a nearly constant behavior under temperature

variations. The worst corner is the FF where the supply current (dissipated power) at T= 100 °C increases from 6.55 nA (11.79 nW) up to 38.7 nA (69.66 nW). On the other hand, the best corner is the SS where the supply current (dissipated power) at T= 100 °C decreases from 6.55 nA (11.79 nW) down to 1.26 nA (2.27 nW). The benefits of the additional transistors can be better appreciated in Fig. 3.66(a) and (b) which provide the static power consumption per bitcell averaged across different process corners under both voltage and temperature variations. Data were extracted from 5k-run Monte Carlo simulations. Fig. 3.66(a) reports the trend of the static power consumption as function of the voltage variations. This figure highlights the effect of using more stacked solutions, especially at high supply voltages. Indeed, at V_{DD}= 1.8 V when moving toward more stacked solutions the static power per bitcell decreases from 175 nW, for the 4T-core, down to 18.9 nW, for the 6T-core, and 3.19 nW, for the 8T-core. On the other hand, at V_{DD}= 0.4 V the static power consumption of all the three solution of Fig. 3.60 approaches to around 250 pW. This can be ascribed to the lower efficiency of the additional transistors at very low voltage operations as stated above. Fig. 3.66(b) illustrates the trend of the static power consumption under temperature variations which proves the effectiveness of the more stacked solutions. Indeed, at T= 0°C passing from a 4T-core based to a 6T-core and 8T-core based solutions decreases the static power from 179 nW down to 24.6 nW and 4.16 nW, respectively.

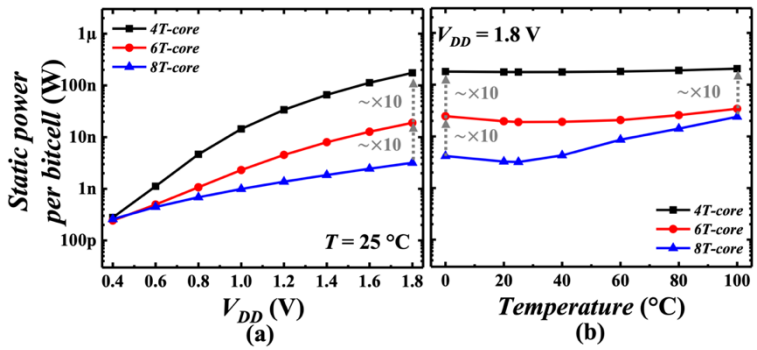


Fig. 3.66. Static power per bitcell averaged over process corners across (a) voltages and (b) temperatures for the three circuits of Fig. 3.60(a)-(c).

On the other hand, at T= 100 °C the static power of 6T-core and 8T-core based solutions is quite similar due to the increase of the overall stability of the 8T solution, as reported in Fig. 3.64(b). Anyway, moving from a 4T-core based toward a more stacked solution (e.g., 6T or 8T) decreases the static power consumption from 204 nW down to around 30 nW.

3.6 Conclusion

In this chapter, we have introduced a novel class of static monostable PUFs based on a sub-threshold voltage divider. The proposed PUF bitcell consists of a sub-threshold voltage divider between two nominally identical sub-circuits as core block along with an output inverter for binarizing the response. More in detail, the mismatch causes one sub-circuit being stronger or weaker than the other thus pushing the output voltage as far from the mid-supply point as low is the voltage sensitivity of the two sub-circuits.

This chapter explores different circuital variant for implementing the sub-circuits with the aim of reducing the native instability. Extensive simulations and analysis on 2T-core, 4T-core, 6T-core,

and 8T-core in 180-nm CMOS technology demonstrate that moving toward more stacked solutions potentially allows suppressing the instability at GK conditions and under voltage variations thus also dramatically reducing the static power consumption per bitcell. This can be achieved at the cost of larger bitcell footprint.

The 4T-core based solution was fabricated in 180-nm CMOS technology with an 8×32-bit array fashion and characterized across V_{DD} ranging from 0.4 – 1.8 V and temperature ranging from 10 – 80 °C. Measurement results showed a uniqueness of 0.493 along with a uniformity of 0.518. The randomness was assessed by a low spatial ACF and by passing all applicable statistical NIST tests. Stability analysis reported a native percentage of unstable bits (BER) of 0.61% (0.13%) at GK conditions along with a sensitivity to voltage and temperature variations of 0.63%/V (0.53%/V) and 0.016%/°C (0.017%/V), respectively thus proving the effectiveness of the proposed solutions.

Chapter 4
PUF implementation in 2D technology

The main contents are extracted from

- **M. Vatalaro,** R. De Rose, M. Lanuzza, G. Iannaccone, and F. Crupi, "*Assessment of 2D-FET Based Digital and Analog Circuits on Paper,*" *Solid-State Electronics*, vol. 185, 2021.
- **M. Vatalaro,** R. De Rose, M. Lanuzza, P. Magnone, S. Conti, G. Iannaccone, and F. Crupi, "*Assessment of Paper Based MoS_2 FET for Physically Unclonable Functions*," *Solid-State Electronics*, vol. 194, 2022.

4.1 Introduction

Two dimensional (2D) materials are nowadays gaining great interest as emerging beyond-CMOS technology. Indeed, their outstanding electrical and mechanical properties make them suitable for transistor electronics [99]-[111]. Moreover, their layered structure makes them suitable for realizing field-effect transistors (FETs) with atomically thin channels. This potentially enables to further scale down the device dimensions while reducing the short channel effects as well as ensuring low field-effect mobility degradation even at low-voltage operations. Indeed, these structures are characterized by strong in-plane bonds and weak covalent perpendicular bonds. The latter allows of exfoliating single atomic layers which can be used as transistor channel for next generation electronics. This results in a higher concentration of carrier density in smaller channel area which reduce the bandgap sensitivity to the channel thickness variations as well as improves the electrostatic of the device. Indeed, thanks to the very thin channels, carriers are confined to less than 1 nm of thickness thus improving the gate electrostatic as well as reducing the short channel effects. Moreover, the weak covalent bonds dramatically reduce the dangling bonds which reduce interface traps, thus improving the device reliability. Ideally, since these materials are used in a FD SOI fashion, they should show some property for being used for electronics purposes, such as:

- ***High enough bandgap*** for ensuring device OFF state, especially for digital applications.
- ***High mobility*** for guarantying faster ON-OFF transitions.

- ***High thermal conductivity***, for dissipating heat faster thus counteracting the FD SOI issues.
- ***Compatibility with Si-CMOS processes***.
- ***Low contact resistances*** for avoiding bottlenecks.
- ***Symmetric channel-oxide interface*** (i.e., similar energy gaps for both electrons and holes) for reducing parasitic currents.
- ***Low density of defects*** in the oxide for preserving the FET performance.

Graphene was the first man-made 2D material, which offers interesting electrical and optical properties combined with good mechanical flexibility. Unfortunately, the absence of an energy bandgap hinders its functionality for digital applications [3]. Among the various 2D semiconducting materials, transition metal dichalcogenides (TMDCs) are premiere candidates for being used as channels for FETs owing to their bandgap tunability [106]. In particular, molybdenum disulfide (MoS_2) is gaining great interest thanks to its large availability in its natural form and simplicity in producing high-quality 2D crystals along with good tunable electrical properties and mechanical flexibility [108]-[111]. The latter property makes them particularly suitable for realizing electronic device on flexible substrates like paper [111]-[114]. Despite paper is still a challenging substrate due to the lack of reliable manufacturing techniques in [111] authors fabricated a paper-based MoS_2 nFET with noticeable transistors performance. This opens the route for next-generation flexible electronics targeting several applications from IoT to hardware security fields (e.g., wearable electronics and smart labels for anti-counterfeiting purposes). During my PhD, I tried to mimic CMOS static designs previous proposed in literature, exploiting experimental measurements from the paper-based MoS_2 fabricated by University of Pisa [111].

4.1.1 Chapter organization

This chapter is organized as follow. Section 4.2 introduces the paper based MoS_2 FET describing the fabrication process and the electrical characterization. Section 4.3 illustrates the adopted LUT based Verilog-A model. Section 4.4 provides PUF circuit schematics along with the simulation results. Finally, Section 4.5 concludes this chapter.

4.2 Fabricated device

4.2.1 Fabrication process

In [111], authors combined chemical vapor deposition (CVD) and inkjet printing through a "channel array" approach. The former was used for the growth of high-quality MoS_2 channel on sapphire substrate. The inkjet printing approach allows designing and fabricating customizable devices and circuits exploiting 2DMs-based inks. These two techniques were combined through a "channel array" technique which consists of transferring strips of CVD-grown MoS_2 onto paper substrate where the other parts of the devices and circuits were customized by the inkjet printing technique. More precisely, source, drain, and gate contacts as well as gate dielectric and connections were fabricated by using specific inks. This approach allows keeping a certain degree

of flexibility and versatility with the high-quality CVD-grown MoS_2 channel is already placed. Fig. 4.1 illustrates the sketch of the fabricated MoS_2 device on paper substrate.

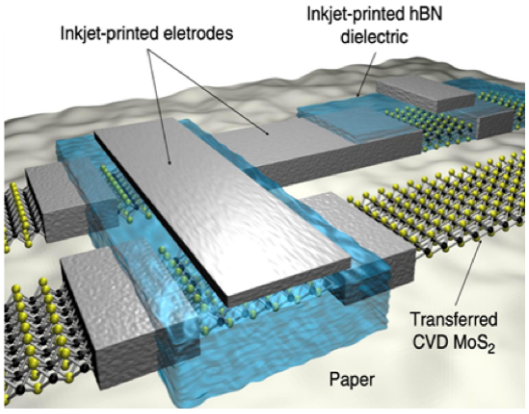


Fig. 4.1. Sketch of fabricated MoS_2 FETs on paper substrate [13].

From this figure, hexagonal boron-nitride (hBN) was used as insulating and printed on the MoS_2 channel to act as gate dielectric while silver was used as drain, gate, and source contacts. More precisely, silver contacts allow reducing the contact resistance, thus creating a low-resistance MoS_2-silver interface at both drain and source sides.

4.2.2 Electrical characterization

Fig. 4.2(a) and (b) illustrate the drain current, I_D, versus drain voltage, V_{DS}, and versus gate voltage, V_{GS}, characteristics, respectively.

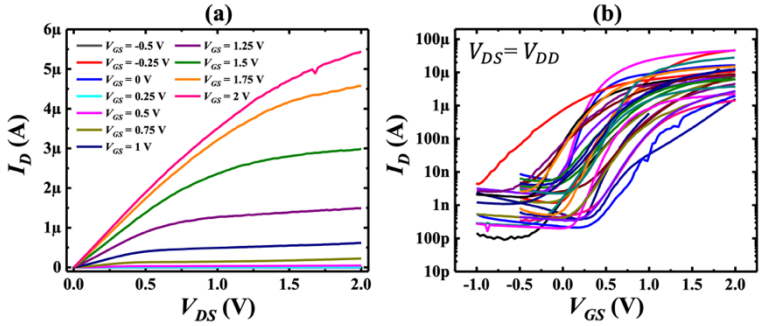


Fig. 4.2. Experimental (a) $I_D - V_{DS}$ characteristics at different V_{GS} for a paper-based MoS_2 with nominal sizing (i.e., L= 80 µm and W= 275 µm) and (b) $I_D - V_{GS}$ characteristics at V_{DS}= V_{DD} for a set of 27 paper-based MoS_2 nFETs from the same manufacturing lot.

In particular, Fig. 4.2(a) shows the I_D vs V_{DS} at different V_{GS} for a paper-based MoS_2 with nominal size (i.e., with channel length, L, of 80 µm and channel width, W, of 275 µm). On the other hand, Fig. 4.2(b) reports the I_D vs V_{GS} at V_{DS}= V_{DD} for a set of 27 paper-based MoS_2 devices from the same manufacturing lot. Fig. 4.3(a) and (b) report the I_D vs V_{DS} in low drain voltage region and the distribution of the $I_{D,ON}/I_{D,OFF}$ ratio (e.g., extracted from the 27 I_D vs V_{DS} characteristics of Fig. 4.2(b)), respectively. Fig. 4.3(a) illustrates the ohmic behavior of the I_D vs V_{DS} characteristics in low-drain voltage region, thus proving the effectiveness of the used silver contacts. Indeed, the

linearity parameter, which describes the relation $I_D \propto V_{DS}^{\gamma}$, showed in the log-log plot of this figure is found to be 1.1 on average.

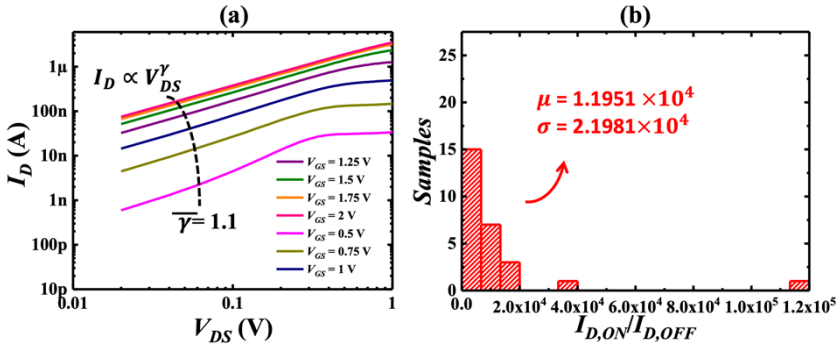


Fig. 4.3. (a) Log-log curves of the I_D vs V_{DS} in low drain voltage region and (b) distribution of the ratio between $I_{D,ON}/I_{D,OFF}$ extracted from I_D vs V_{GS} characteristics of 27 devices of Fig. 4.2(b).

This indicates a good contact between the CVD MoS_2 and the inkjet-printed silver electrodes [111]. Fig. 4.3(b) report the $I_{D,ON}/I_{D,OFF}$ distributions of the 27 I_D vs V_{GS} characteristics reported in Fig. 4.2(b). From this distribution the average value is $\sim 1.2\times 10^4$ which demonstrates a good device electrostatic. For each tested device was extracted the threshold voltage (V_{TH}) and the field-effect mobility (μ_{FE}) from the linear fitting of the square root of the I_D (i.e., $\sqrt{I_D}$) vs V_{GS} curve in saturation regime (i.e., $V_{DS} > V_{GS} - V_{TH}$) [111], [115] and [116]. Fig. 4.4 illustrates the V_{TH} and μ_{FE} extraction procedure for a representative device.

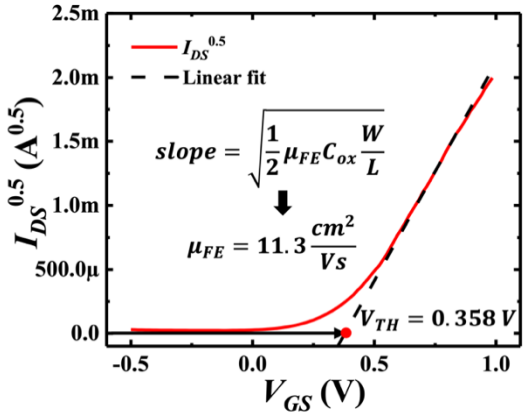


Fig. 4.4. Extraction of the threshold voltage (V_{TH}) and field-effect mobility (μ_{FE}) from $\sqrt{I_D}$ vs V_{GS} curve at $V_{DS} = V_{DD}$ for a representative device.

More precisely, the V_{TH} was extrapolated as the x-axis intercept of the fitting straight line passing through the maximum transconductance point, whereas the μ_{FE} was estimated considering the current equation in saturation region which is given by

$$I_D = \frac{1}{2}\mu_{FE}C_{OX}\frac{W}{L}(V_{GS} - V_{TH})^2 \qquad (4.1)$$

where the oxide capacitance per unit area C_{OX} was set to 230 nF/cm^2, i.e., the average value as extracted from measurements on parallel plate capacitor structures [111]. From (4.1) the square root of the current is given by

$$\sqrt{I_D} = \sqrt{\frac{\mu_{FE} C_{OX}}{2} \frac{W}{L}} (V_{GS} - V_{TH}) \quad (4.2)$$

From (4.2) the field-effect mobility can be extracted as follow.

$$\mu_{FE} = 2 \frac{L}{W} \frac{1}{C_{OX}} \left(\frac{\partial \sqrt{I_D}}{\partial V_{GS}} \right)^2 \quad (4.3)$$

For the representative device the extracted values of V_{TH} and μ_{FE} are 0.358 V and 11.3 $cm^2V^{-1}s^{-1}$, respectively. As a result, Fig. 4.5(a)-(b) provide the extracted statistical distributions extracted from the curves of Fig. 4.2(b) along with the correlation analysis, respectively. Fig. 4.5(a) shows the statistical distribution of the extracted threshold voltage (V_{TH}) which is well fitted by a normal distribution. From this figure the average V_{TH} value is of 0.387 V with a standard deviation of 0.357 V (e.g., with a variability of 92.25%). On the other hand, Fig. 4.5(b) reports the statistical distribution of the extracted field-effect mobility (μ_{FE}). Differently from the threshold voltage distribution, this is well fitted by an exponential distribution. From this distribution the mean value is equal to the standard deviation ($\mu = \sigma$) equals to 8.35 $cm^2V^{-1}s^{-1}$. These two distributions report a very high variability showed by these paper-based devices.

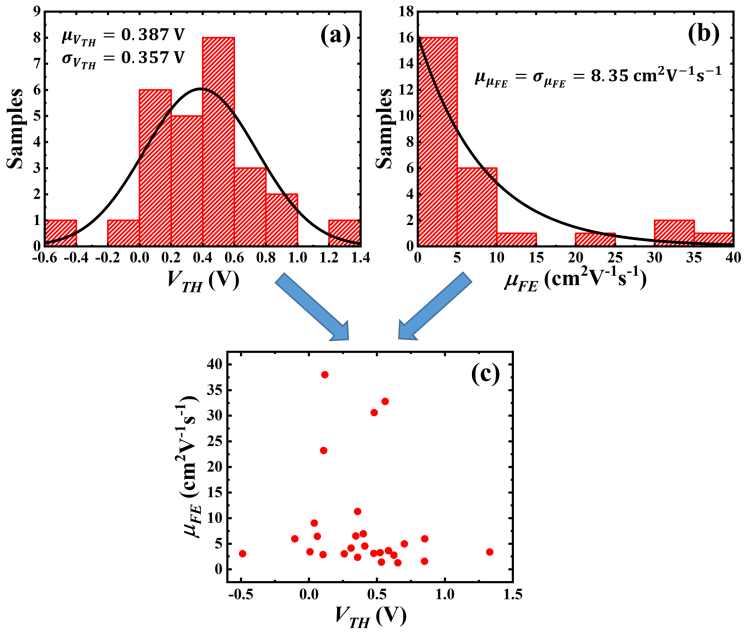


Fig. 4.5. Statistical distribution of (a) threshold voltage (V_{TH}), and (b) field-effect mobility (μ_{FE}). Finally, (c) $\mu_{FE} - V_{TH}$ scatter plot.

Finally, Fig. 4.5(c) report the $\mu_{FE} - V_{TH}$ scatter plot, which highlights that no correlation exists between these two extracted parameters.

4.3 Simulation framework

To perform and assess the paper-based MoS_2 FET performance for being used for PUF applications a simulation framework was built in order to perform circuit simulations within cadence virtuoso. Fig. 4.6 summarized the implemented simulation framework.

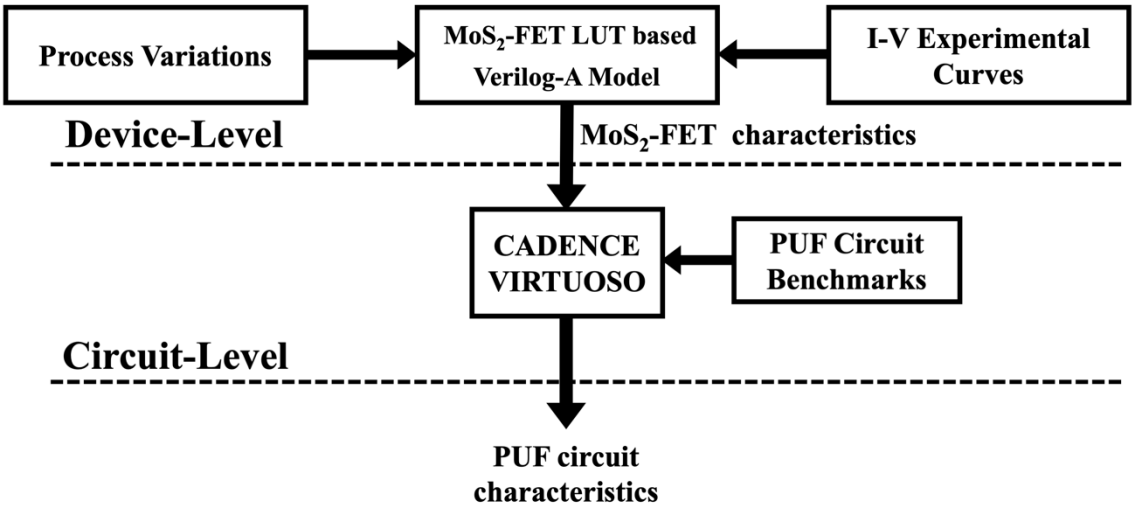


Fig. 4.6. The adopted simulation framework.

From this figure, a LUT-based Verilog-A model was built and then imported in cadence virtuoso environment for enabling circuit simulations. In particular, experimental data were exploited to setup a LUT-based model of a 3-terminal device [116]-[117]. More precisely, the experimental I_D versus V_{DS} curves showed in Fig. 4.2(a) were used for generating a look up table (LUT) with the nominal I_D vs V_{DS} characteristic (i.e., referring to a device with nominal sizing with 80 µm as channel length and 275 µm as channel width). Moreover, to enable tuning of transistor strength, the developed model includes a sizing parameter (i.e., k_M) representing a multiplying factor for the effective width (i.e., $W_{eff} = W_{nom} \times k_M$, with $W_{nom} = 275\ \mu m$) and hence for the current as summarized in Fig. 4.7(a). The model also takes into account the effect of process variations in terms of V_{TH} and μ_{FE} variability according to the distributions reported in Fig. 4.5(a) and (b). Moreover, the effect of different sizing (i.e., different k_M) on both V_{TH} and μ_{FE} variability is considered according to the well-known Pelgrom's law [24]. From Fig. 4.7(b), the threshold voltage variability was modeled through a normal distribution whose probability distribution is given by

$$PDF(x, \mu_{V_{TH}}, \sigma_{V_{TH}}) = \frac{1}{\sigma \cdot 2\pi} e^{-\frac{1}{2}\left(\frac{x-\mu_{V_{TH}}}{\sigma_{V_{TH}}}\right)^2} \quad (4.4)$$

Where $\mu_{V_{TH}}$ represents the threshold voltage main value equals to that reported in Fig. 4.5(a) (i.e., 0.387 V), whereas the standard deviation $\sigma_{V_{TH}}$ is equal to $\sigma_{nom,V_{TH}}/\sqrt{k_M}$ with $\sigma_{nom,V_{TH}}$ equals to 0.357 V as reported in Fig. 4.5(a). This allows scaling the threshold voltage standard deviation as $\sigma_{V_{TH}} \propto 1/\sqrt{W_{eff}L}$. In this way, in the nominal case (i.e., k_M= 1) the threshold voltage standard deviation is equal to its nominal value of 0.357 V (i.e., $\sigma_{V_{TH}} = \sigma_{nom,V_{TH}} = 0.357$ V) as reported

in Fig. 4.7(b). From [24] the field-effect mobility (μ_{FE}) variability is expected to scale as $\sigma_{\mu_{FE}}/\mu_{\mu_{FE}} \propto 1/\sqrt{W_{eff}L}$. Therefore, assuming the mean value of the field-effect mobility independent of sizing, its standard deviation also scales as $\sigma_{\mu_{FE}} \propto 1/\sqrt{W_{eff}L}$, i.e., $\sigma_{\mu_{FE}} \propto \sigma_{nom,\mu_{FE}}/\sqrt{k_M}$ with $\sigma_{nom,\mu_{FE}}$ being the standard deviation nominal value reported in Fig. 4.5(b) (i.e., 8.35 $cm^2V^{-1}s^{-1}$, which corresponds at k_M= 1).

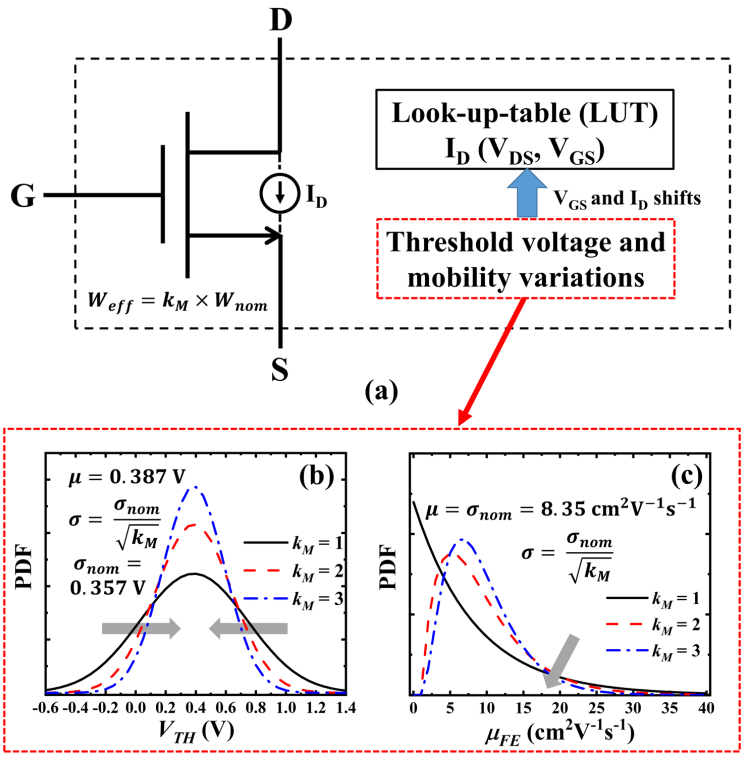


Fig. 4.7. (a) Sketch of the LUT-based Verilog-A model used for the 3-terminal device representing the paper-based MoS_2 nFET. (b) Modeling of the threshold voltage (V_{TH}) variability through a normal distribution and (c) modeling of the field-effect mobility (μ_{FE}) variability using an Erlang distribution.

From Fig. 4.7(c), for enabling μ_{FE} scaling with the transistor sizing, Erlang distribution was used whose probability distribution is given by

$$PDF(x, k_M, \theta) = \frac{x^{k_M-1}e^{-\frac{x}{\theta}}}{\theta^{k_M}(k_M-1)!} \qquad (4.5)$$

where k_M and θ are respectively the shape and the scale parameters, which determine $\mu_{\mu_{FE}} = k_M\theta$ and $\sigma_{\mu_{FE}} = \sqrt{k_M}\theta$. From (4.5), when k_M= 1 we obtain an exponential distribution (i.e., $\sigma_{\mu_{FE}} = \mu_{\mu_{FE}} =$ 8.35 $cm^2V^{-1}s^{-1}$), thus corresponding to the fitting distribution extracted from experimental data of Fig. 4.2(b), as showed in Fig. 4.7(c). On the other hand, when increasing k_M (i.e., $k_M > 1$) while assuming the mean value, $\mu_{\mu_{FE}}$, independent from the transistor sizing and hence k_M, the scale parameter scales as $\theta = \mu_{\mu_{FE}}/k_M = \sigma_{\mu_{FE}}/k_M$. Accordingly, the field-effect mobility standard deviation scales as $\sigma_{\mu_{FE}} = \sigma_{\mu_{FE}}/\sqrt{k_M}$, in agreement with what discussed above and showed in Fig. 4.7(c). As result, when performing Monte Carlo simulations, the developed Verilog-A code translates a random deviation of V_{TH} and μ_{FE} from their nominal values into corresponding V_{GS} and I_D shifts, respectively, on the tabulated current-voltage characteristics. It

is also worth pointing out that random samples of V_{TH} and μ_{FE} are generated assuming no statistical correlation, as evidenced by statistical measurement showed in Fig. 4.5(a)-(c).

4.4 PUF circuit benchmarks

The model described above was built for enabling PUF circuit simulations. In particular, different already proposed PUF solutions were tested using the experimental measurements of these paper-based MoS_2 devices. For safe comparison, each tested PUF circuit was tested with the same bitcell fashion sketched in Fig. 4.8. From this figure, the bitcell consists of a core block, which translates the process variations into a random V_X voltage, along with an output buffer for ensuring high enough impedance at V_X node as well as for generating the output bit. The latter, in particular, was implemented using two resistor transistor logic (RTL) inverter showed in the bottom right side of Fig. 4.8.

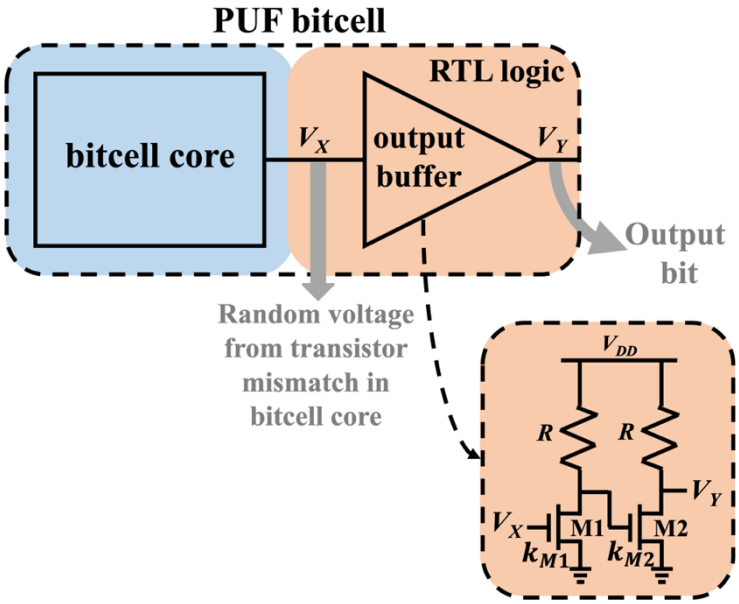


Fig.4.8. Conceptual diagram of simulated PUF bitcell with the schematic of the output buffer.

4.4.1 RTL Inverter design

The adoption of an output buffer instead of a single inverter is mainly due to the low gain showed by the RTL inverter, where the resistor R can be effectively implemented by inkjet-printed graphene. Finally, the output V_Y voltage, thus represents the output bit of the bitcell. The use of the RTL logic inevitably degrades the performance of the inverter when compared to the conventional CMOS solutions. Indeed, a proper inverter sizing is required for reducing the sensitivity to the noisy conditions. Fig. 4.9(a)-(d) summarize the simulation results obtained for the RTL inverter at different resistor (R) and multiplying factor (k_M) values without considering the process variations [16].

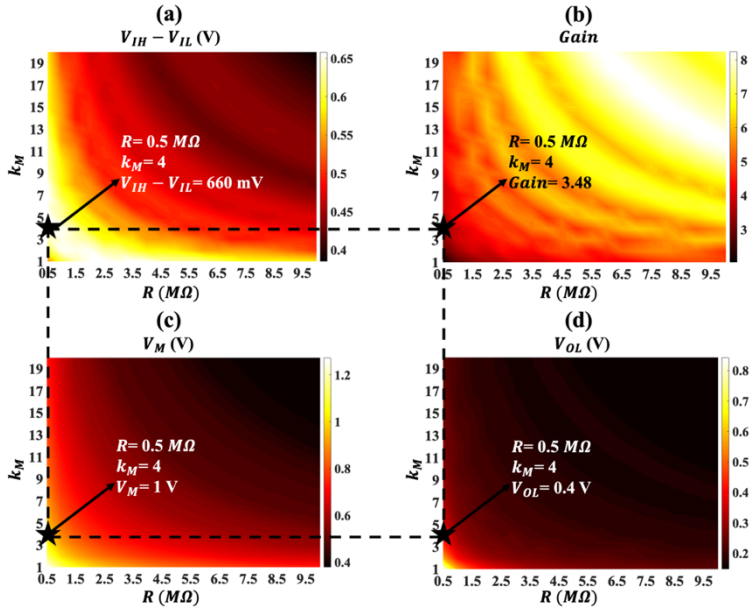


Fig. 4.9. Simulation results of the RTL inverter at V_{DD}= 2 V and T= 25 °C: (a) unstable input region ($V_{IH} - V_{IL}$), (b) output gain, (c) logic threshold (V_M), and (d) maximum low output voltage (V_{OL}).

Fig. 4.9(a) report the color map of the unstable input region (i.e., $V_{IH} - V_{IL}$) as function of the resistor value (i.e., R) and the effective transistor channel width (i.e., $W_{eff} = k_M \times W_{nom}$ with W_{nom} = 275 µm). From this figure the amplitude of this region decreases when increasing both resistor value and multiplying factor. Fig. 4.9(b) illustrates the gain trend as function of the inverter sizing. This trend is strictly related to the map of the unstable region. Indeed, higher gain will result in a narrower unstable input region since the values of V_{IH} and V_{IL} comes closer to the inverter logic threshold (ideally the mid-supply point). Indeed, when increasing both R and k_M values the gain (unstable region) also increases (decreases). This can be ascribed to the fact that M1 translates input voltage variations (i.e., ΔV_{IN}) into drain current variations (i.e., $g_m \Delta V_{IN}$) and hence into output voltage variations (i.e., $-g_m R \Delta V_{IN}$). This means that, for a given ΔV_{IN} higher values of R and k_M result in a higher ΔV_{OUT}. From Fig. 4.9(c), the logic threshold (i.e., V_M) decreases when increasing the inverter sizing. This is mainly due to the fact that the increase of both resistor and multiplying factor values increases the strength ratio between the pull-down network (PDN) and the pull-up network (PUN) thus resulting in a HIGH-LOW transition before the mid-supply point ($V_{DD}/2$). Similarly, from Fig. 4.9(d) the maximum low output voltage (i.e., V_{OL}) decreases when increasing both R and k_M. This can be also ascribed to the higher PDN strength compared to the PUN which results in a lower V_{OL} (e.g., for a given output voltage the PDN requires lower voltage drop compared to the PUN). For our purposes the output inverter was sized with R= 0.5 MΩ and k_M= 4 for achieving a logic threshold equals to the mid-supply point for guarantying a proper randomness at the output of the PUF bitcell, as pointed out by Fig. 4.9. Fig. 4.10 reports the input-output transfer characteristic of the inverter along with the voltage gain for the chosen sizing at nominal conditions (i.e.,V_{DD}= 2 V and T= 25 °C). From this figure the adoption of R= 0.5 MΩ and k_M= 4 allows achieving a logic threshold (i.e., the input voltage for which $V_{IN} = V_{OUT}$) equals to 1 V.

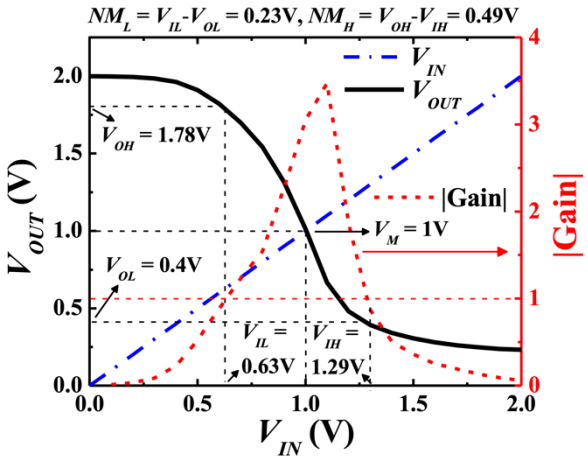


Fig. 4.10. Input-output transfer characteristic and voltage gain for R = 0.5 MΩ and k_M = 4 at V_{DD} = 2 V and T= 25 °C.

However, this is achieved at the cost of higher maximum low output voltage (V_{OL}= 0.4 V), and hence narrower voltage swing (i.e., $V_{OH} - V_{OL}$= 1.38 V), which results in a low noise margin (i.e., $NM_L = V_{IL} - V_{OL}$) of 0.23 V with a high noise margin (i.e., $NM_H = V_{OH} - V_{IH}$) of 0.49 V. Moreover, the chosen sizing results in a larger input unstable region ($V_{IH} - V_{IL}$= 660 mV) and hence in a low voltage gain (i.e., 3.48) [16].

4.4.2 Analyzed PUF solutions

Concerning the bitcell cores, Fig. 4.11(a)-(d) illustrate four different static monostable designs considered in this thesis along with the transistor/resistor sizing. More precisely, Fig. 4.11(a) mimics the concept proposed in [49], where two back-to-back current mirrors translate transistors mismatch into voltage variations. Due to the lack of paper-based MoS_2 pFET, the p-version of the current mirror was replaced by a pair of resistors. These resistors were sized for achieving a V_X voltage equals to the mid-supply point at nominal conditions (i.e., when disabling the transistors mismatch). As reported in Fig. 4.11(a), for ensuring an adequate randomness (i.e., the same probability of having a bit '0' or '1' at the output) the adopted resistors should be equal to 500 $M\Omega$. On the other hand, minimum transistor sizing (i.e., with the multiplying factors, k_M, equal to the minimum value of 1) was used for maximizing the transistor mismatch and hence the V_X voltage spread. Fig. 4.11(b) refers to the solution proposed in [56], where the PUF response is generated by amplifying the threshold voltage difference between two inverting logic gates, i.e., in this case two RTL NAND gates with an input enable signal V_{EN}. As in the previous circuit, the transistors sizing in equal to the minimum one (i.e., k_M= 1) for maximizing the transistors variability. The PUN was composed by a resistor whose value was chosen for having $V_{DD}/2$ as output at nominal conditions (i.e., when disabling the process variations). As reported in Fig. 4.11(b) 6.8 $M\Omega$ resistor ensure an adequate randomness at the output.

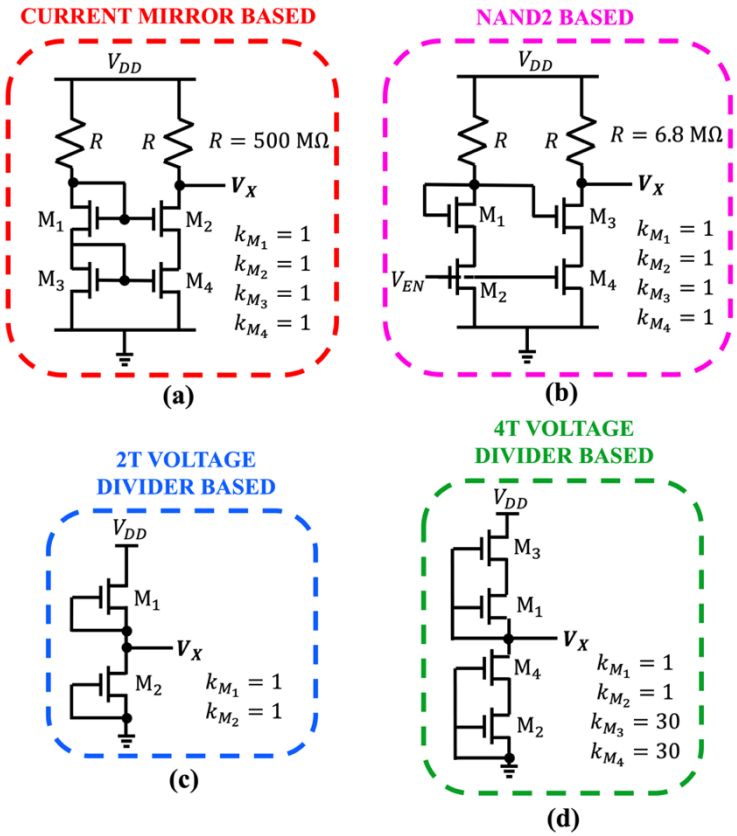


Fig. 4.11. Schematic of the implemented PUF bitcell cores along with the transistor/resistor sizing: (a) current mirror based, (b) NAND2 based, (c) 2T voltage divider, and (d) 4T voltage divider.

Fig. 4.11(c) refers to the solution proposed in [97], composed by the 2T sub-threshold voltage divider described in the sub-section 3.1. Differently, from the circuit described in sub-section 3.1 a nFET-based version was proposed here due to the lack of pFETs. More precisely, the core circuit consists of a voltage divider between two zero-V_{GS} nFETs where due to the mismatch the V_X voltage differs from $V_{DD}/2$ as low is the DIBL effect of the two FETs. Indeed, the DIBL effect plays a key role on the mismatch amplification thus resulting in trade-off between long channel devices which ensure lower DIBL effect as well as lower variability and long channel devices which ensure higher DIBL effect but at the same time higher variability. Here, not being able to act on the channel length of the device within the adopted modeling, the two nFETs were minimum sized (i.e., k_M= 1) for increasing the variability. Similar to what said in the previous chapter, the sub-threshold operation ensures higher variability, due to the exponential relationship between the current and the threshold voltage, as well as lower power consumption. Moreover, the use of only pFET allows avoiding the adoption of the resistors. As result, in absence of mismatch, the V_X voltage is equal to the mid-supply point ($V_{DD}/2$) while when the mismatch occurs M1 and M2 translates process variations into V_X deviations from the mid-supply point. Finally, Fig. 4.11(d) refers to the solution proposed in this thesis [13] which consists of a 4T sub-threshold voltage divider between two nominally identical sub-circuits. More precisely the circuits implement the nFET-version of the solution described in chapter 3 which includes two zero-V_{GS} nFETs (i.e., M1 and M2) which act as main mismatch source and two negative-V_{GS} nFETs (i.e., M3 and M4) which act as mismatch booster. Similar to the previous solution, the nFET-only implementation avoids of using the resistors, whereas concerning the transistor sizing,

the guidelines described in the previous chapter were used. Indeed, M1 and M2 were sized with minimum sizing (i.e., k_M= 1) for maximizing their variability and hence their strength difference, whereas larger sizing was used (i.e., k_M= 30) for M3 and M4 for ensuring an adequate voltage drop across M1 and M2 and hence for maximizing the effect of transistors mismatch on the V_X voltage variability. As result, M1 and M2 translate the transistor mismatch into a voltage difference between their voltage drops (i.e., $V_{DS,1} - V_{DS,2}$) and M3 and M4 translate such voltage difference into a high V_X deviation from the mid-supply point.

4.4.3 Simulation results

The circuits described above were tested at nominal conditions (i.e., V_{DD}= 2.0 V and T= 25 °C) from 5k-run Monte Carlo simulations without including the variability of the resistors. Fig. 4.12 (a)-(h) report the statistical distributions of the V_X and V_Y voltages for each mimicked bitcell under process variations.

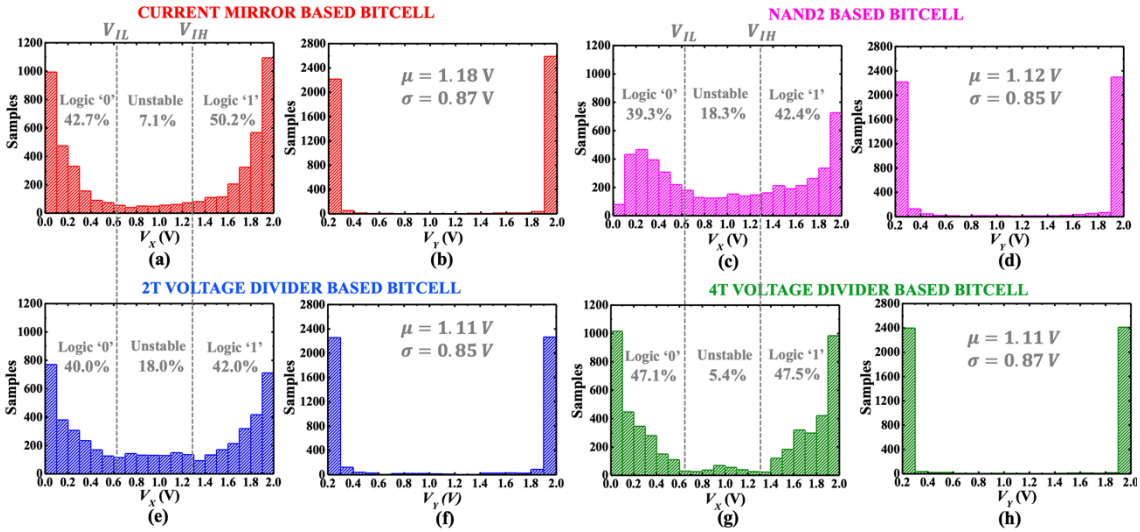


Fig. 4.12. Statistical distributions of the voltages V_X and V_Y as provided by the bitcell core under process variations from 5k-run Monte Carlo simulations at nominal conditions (i.e., V_{DD}= 2.0 V and T = 25 °C): (a) current mirror based bitcell, (b) NAND2 based bitcell, (c) 2T sub-threshold based bitcell, and (d) 4T sub-threshold based bitcell.

The V_X histograms of Fig. 4.12 also report the estimated breakdown among logic '0', logic '1' and potential unstable bits, defined as the V_X voltage samples that fall in the unstable region of the subsequent inverter, defined as the difference between the minimum high and maximum low input voltages (i.e., $V_{IH} - V_{IL}$) extracted from the nominal input-output characteristic of the inverter showed in Fig. 4.10, thus resulting into uncertain bits under time-varying sources of variation such as voltage, temperature and noise. Fig. 4.12(a) and (b) show the V_X and V_Y distributions of the current mirror based solution of Fig. 4.11(a). Such design provides a V_X distribution with high deviation from the mid-supply point as demonstrated by 7.1% of potential unstable bits. However, the use of resistors slightly affects the randomness, thus resulting in an unbalanced percentage of logic '1' and '0' at the output (i.e., 50.2% and 42.7%, respectively). This is also appreciable by observing the V_Y distribution of Fig. 4.12(b). Fig.12(c) and (d) provide the V_X and V_Y distributions of the NAND2 based solution. In particular, from Fig. 4.12(c) the percentage of potential unstable bits is of 18.3% thus highlighting that the gain provided by the

two NAND gates is not large enough for pushing the V_X samples far from $V_{DD}/2$. Moreover, also here the use of resistors for compensating the lack of pFETs slightly affect the randomness thus resulting in an unbalanced percentage of logic '1' and '0' (i.e., 42.4% and 39.3%, respectively). Moreover, the low voltage gain provided by the subsequent buffer results in a certain percentage of non-well-defined bits (i.e., bits that are close to the mid-supply point even after two conversion stages) as shown in Fig. 4.12(d). Fig. 4.12(e) and (f) report the V_X and V_Y distributions referring to the 2T sub-threshold voltage divider of Fig. 4.11(c). From 4.12(e), despite the high variability showed by the device the V_X distribution highlights a high percentage of potential unstable bits (e.g., 18.0%) thus indicating that these devices show a high DIBL effect. However, the randomness was ensured by the adoption of two nominally identical sub-circuits which guarantees a balanced percentage of logic '1' and '0' (i.e., which are 42% and 40%, respectively). This can be also observed in the V_Y distribution of Fig. 4.12(f). Finally, Fig. 4.12(g) and (h) report the V_X and V_Y distributions for the 4T sub-threshold voltage divider of Fig. 4.11(d). From Fig. 4.12(g), similarly to what seen with the 2T core solution, the adoption of two nominally identical sub-circuits ensures a good randomness. Indeed, the percentages of logic '1' and '0' are 47.5% and 47.1%, respectively. Moreover, the addition of the reverse biased FETs (i.e., M3 and M4) leads to higher V_X voltage spread, thus resulting in a dramatic decrease of the potential unstable bits (i.e., 5.4%). This can be better appreciated in Fig. 4.12(h) where, despite the low voltage gain provided by the output buffer, the most part of V_Y samples fall close to the two edges (i.e., V_{DD} or 0.2 V). It is important to stress that the deviation from the mid-supply point of the average values showed in the V_Y distributions of Fig. 4.12 is related to the fact that the adopted RTL buffer shows a logic '0' higher than 0 V as you can observe in Fig. 4.10. Finally, Table IV compares the obtained results for the simulated PUF solutions.

TABLE IV. SUMMARY RESULTS FOR PUF DESIGNS OF FIG. 4.11

PUF bitcell	Unstable bits (%)	$P_r(1)$	Resistor count	FET count	FET total area (mm^2)	Core static power (nW)
Current mirror	7.1	0.538	4	6	0.264	8.86
NAND2	18.3	0.508	4	6	0.264	572.6
2T voltage divider	18.0	0.498	2	4	0.22	2.46
4T voltage divider	5.4	0.498	2	6	1.54	0.91

More precisely, the Table compares some common PUF metrics, such as the percentage of unstable bits (i.e., the percentage of V_X samples that fall in the unstable region of the inverter, $V_{IH} - V_{IL}$), the probability of generating a bit '1' (i.e., $P_r(1)$) at the output of the bitcell (i.e., the percentage of V_X samples above than the inverter logic threshold, V_M) which directly estimates the randomness, the number of adopted resistors and FETs (including that used for implementing the RTL buffer), the total area occupation only referred to transistors (i.e., neglecting the contribution associated with resistors) approximately evaluated as $W_{eff} \times L$ for each FET, and the static power consumption associated with the core circuit. According to data discussed above the 4T voltage divider-based cell shows a lowest percentage of unstable bits equals to 5.4%. Indeed, this circuit shows an improvement of $1.3-3.4\times$ as compared to other solutions, while also ensuring a randomness close to the ideal value of 0.5. Moreover, the 4T core circuit also shows the lowest static power consumption equals to 0.91 nW. This circuit shows an improvement of $2.7-629.2\times$ as compared to other solutions. This is achieved at the cost of larger bitcell area.

Indeed, the value of 1.54 mm^2 is 5.8−7 × higher than the other solutions due to the need of having large M3 and M4 FETs within the core circuit as discussed above.

4.5 Conclusion

In this chapter we have investigated static PUF solutions using novel paper based MoS_2 devices. Circuit simulations were performed by using a LUT-based Verilog-A model to describe the characteristics of 2D-FETs into Cadence Virtuoso environment.

The developed model was calibrated with experimental measurements of MoS_2 FETs fabricated on paper substrate combining chemical vapor deposition and inkjet printing approach using a channel array technique.

Data from 27 devices from the same manufacturing lot show interesting performance along with a high variability. Indeed, the extracted distributions of both threshold voltage and field-effect mobility show a mean value of 0.387 V and 8.35 $cm^2/(V \cdot s)$ with a standard deviation of 0.357 V and 8.35 $cm^2/(V \cdot s)$, respectively. These data were exploited for accounting the effect of process variability within the model.

Four PUF bitcell implementations were analyzed, which use different topologies for the core block. Among the investigated PUF circuits, the solution based on the 4T voltage divider showed the lowest percentage of unstable bits at the cost of larger bitcell area. Anyway, simulation results demonstrate that the large variability exhibited by these paper-based devices can be effectively exploited for implementing cryptographic primitives such as PUFs. This makes this emerging technology suitable for applications of flexible electronics targeting the field of hardware security (e.g., anti-counterfeiting smart labels).

Chapter 5
PUF-based smart tag

The main contents are extracted from

- A. Rullo, C. Felicetti, **M. Vatalaro,** R. De Rose, M. Lanuzza, F. Crupi, and D. Saccà, "*PUF-Based Authentication-Oriented Architecture for Identification Tags," Submitted to IEEE Transactions on Dependable and Secure Computing*.

5.1 Introduction

The increase of the number of connected devices and hence of the ubiquitous objects to the Internet, rapidly pushes the demand of preserving the information down to the edge nodes. While improving our life, the advent of the IoT scenario brings several challenges to face. Indeed, along with the number of connected devices, the number of potential security threats is also increasing in the same manner. Among the various threat topologies, counterfeiting is becoming a serious problem which affects producers and customers all over the world [118]. Moreover, globalization and e-commerce heavily amplified this problem generating economic problems for companies. In the last years, the counterfeiting issue is requiring ad-hoc strategies for being counteracted since it becomes crucial in some critical domains such as military, food, and medicine where the counterfeited parts can endanger the human health. Indeed, some counterfeited products could show poor quality and not meet the minimum safety and security standards [118]. Reverse engineering techniques make possible to create identical copies which are sold as authentic in order to deceive the final customer and/or regulation authorities. Moreover, different parties are involved during the IC fabrication thus making the counterfeiting problem ever more crucial. Generally, for counteracting this issue it is necessary to integrate platforms capable of recording every step among the involved companies along the entire supply chain. From one side, these platforms allow of providing to the involved actors precise information on what third companies are doing along the chain to make products while also ensuring that recorded transactions are truthful and not tempered. However, the integration of these platforms is not an easy task and could represent a single point of failure or a performance bottleneck.

5.1.1 Previous blockchain based approaches

Blockchain [119] is an emerging technology which can effectively represent a solution for this class of threats. In short, blockchain is defined as a digital register whose entries are grouped into blocks, concatenated in chronological order, and whose integrity is guaranteed by the use of cryptographic algorithms. Many groups combine blockchain solutions along with near-field

communication (NFC) [120] and radio-frequency identification (RFID) [121] tags for preventing counterfeits in the post supply chain. However, RFID-based approaches suffer from cloning attacks thus making this solution not suitable for anti-counterfeiting purposes. A possible solution consists of combining the blockchain-based supply chain along with the physically unclonable functions (PUFs) as anti-counterfeiting element, thus resolving the issues related to the RFID-based approach.

5.1.2 Chapter organization

The chapter is organized as follow. Section 5.2 describes the proposed smart tag architecture and its operative principle. Section 5.3 provides a description of the TRNG module. Section 5.4 describes the ECC component. Section 5.5 illustrate the I/O functionality. Section 5.6 provides a security analysis. Finally, Section 5.7 concludes this chapter.

5.2 Smart tag architecture

Here, a PUF-based smart tag, as solution for anti-counterfeiting applications, will be described. As described in the previous chapters, silicon PUFs are devices which exploit random process variations for generating a unique ID for each device like a digital fingerprint. The use of these devices as digital ID has recently attracted great attention due to their inherent interesting cryptographic-oriented properties such as uniqueness, randomness, unclonability and unpredictability. Moreover, leveraging on process variations PUFs represent low-cost and more secure solution compared to the conventional NVM-based approach. Since identification tags are considered resource-constrained devices, PUFs are considered promising solutions for being used as security primitives. However, the PUF reproducibility can be affected by noise and/or different environmental conditions thus requiring fault tolerance techniques such as the fuzzy extractor.
However, these techniques could represent a bottleneck in the production process and offers an attack vector to an adversary who can use the recovery data to reduce the complexity in guessing the secret key. Moreover, authentication protocols should be reliable at each attack topology, for example, the memory usage must be minimized since it represents a vulnerability point of the system. Indeed, the authentication protocol must limit the exchange of secret information with other entities for preventing the analysis of tag communication which could leads to the disclosure of that secrets. In few words, the authentication protocols must be designed for avoiding that secret data need to be stored by other entities. Another point is that the use of true random number generator (TRNG) instead of the pseudo random number generator (PRNG) increase the security due to their vulnerable algorithmic nature. Moreover, considering the applications for which there are intended it is important to make the tag computations energy efficient. Here, we propose a PUF-based, memory-less, authentication-oriented integrated circuit for identifications tag which implements an elliptic curve cryptography (ECC)-based, one-way authentication protocol which includes: (*i*) the highly stable monostable PUF circuit described in chapter 3, (*ii*) an ECC-based one-way authentication protocol which exploits the PUF for performing secure authentication task and avoid of storing secret data at the verifier side, (*iii*) a memory-less, authentication oriented tag architecture which features that the private and public keys are generated by a true

source of randomness and the private key is not stored in any-memory and hence generated dynamically when required. Moreover, the produced output is an ECDSA digital signature used as digital fingerprint for the IoT device. Fig. 5.1 illustrates the high-level tag architecture.

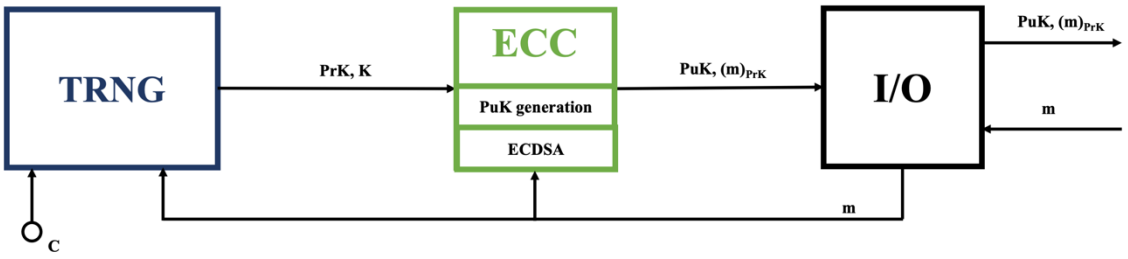


Fig. 5.1. High-level tag architecture.

From Fig. 5.1 the proposed tag architecture consists of three ad-hoc designed components including a TRNG module, ECC module and an Input/Output interface and does not include memory elements. The authentication protocol works as follow: the verifier submits a message m of 256 bits (i.e., the hash of a challenging input) to the tag; the tag dynamically generates the private key (i.e., PrK) and a nonce (i.e., k) by means of its embedded deterministic TRNG component and uses them for generating a digital signature, by means of the ECDSA algorithm, for m (i.e., $(m)_{PrK}$) and the public key (i.e., PuK); finally the pair composed by the signed message ($(m)_{PrK}$) and public key (i.e., PuK), $< (m)_{PrK}, PuK >$, is given back to the verifier which authenticates the tag by simply verifying the signature using its public key. Fig. 5.2 offers a low-level view of the proposed tag architecture.

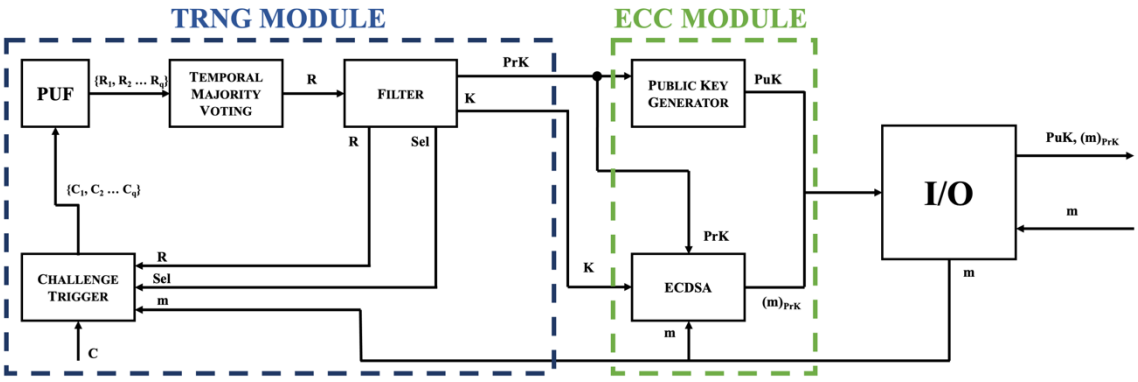


Fig. 5.2. Low-level tag architecture.

From Fig. 5.2, the TRNG module is composed by the static monostable PUF described in chapter 3 along with a challenge trigger (CT), the temporal majority voting (TMV) component and the filter. On the other hand, the ECC module consists of the public key generator and the elliptic curve digital signature algorithm (ECDSA).

5.3 TRNG module

Within the TRNG module, the PUF circuit generates a random private key and a nonce when stimulated by an input challenge, when they are required for the authentication. The challenge is

selected by the challenge trigger which dynamically switches between different input according to the required task (i.e., private key or nonce). The temporal majority voting implements a temporal averaging technique cancelling out the unstable bits. It implements an error masking mechanism used to achieve the required response stability. Finally, a filter ensures that the generated output number (i.e., the private key and nonce) meet the ECDSA requirement of falling into the range [1, n-1] where n is the order of the elliptic curve. More precisely, these components interact as follow: when the device receives a message m through the I/O component, the CT module enables a hardwired challenge of 256-bit to be given as input of the PUF component. This challenge is generated q times so that the PUF generates q responses to be temporal filtered by the TMV. The cleaned response R is given to the filter component for checking if it satisfies the ECDSA requirements. If not, the previous steps are repeated until the generated response satisfies the ECDSA requirements. Finally, the nonce k is generated following the same procedure used for generating the private key with the only difference that the CT module stimulates the PUF instance using m instead of C.

5.3.1 PUF module

The PUF module consists of the circuit published in [13] and described in Sub-Chapter 3.2. Fig. 5.3(a)-(c) illustrate the bitcell design concept along with the statistical distributions of the output voltages of both bitcell core (i.e., V_X) and inverter (i.e., V_Y).

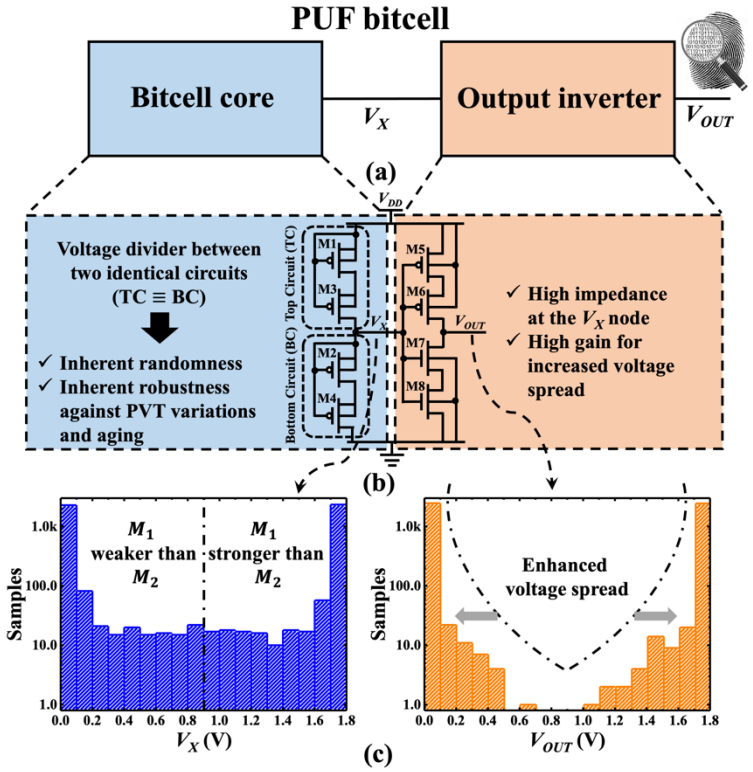


Fig. 5.3. (a) Block-level and (b) transistor-level views of the proposed bitcell. (c) Statistical distributions of V_X and V_Y voltages from 5k-run Monte Carlo simulations at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C).

Briefly, the proposed PUF consists of a 4T sub-threshold voltage divider between two nominally identical sub-circuit (i.e., top circuit, TC, and bottom circuit, BC, with nominally TC ≡ BC), as showed in Fig. 5.3(b). Each sub-circuit is composed by one zero-V_{SG} transistor which acts as main variability source (i.e., M1 for the TC and M2 for the BC) and one negative-V_{SG} transistor which acts as variability booster (i.e., M3 for the TC and M4 for the BC). As result, in absence of mismatch the V_X voltage is equal to the mid-supply point (i.e., $V_{DD}/2$). When mismatch occurs, M1 and M2 translate their mismatch into a difference between their voltage drops (i.e., $V_{SD,1} - V_{SD,2}$). Then, M3 and M4 translate such voltage difference in a V_X deviation from the mid-supply point. The PUF circuit was designed in 180-nm CMOS technology. Fig. 5.3(c) provides the statistical V_X distribution generated by the bitcell core from 5k-run Monte Carlo simulation at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C). The figure also reports the statistical V_Y distribution generated by the output inverter. From this figure is notable how the most part of V_X samples are pushed very close to the two edges (i.e., V_{DD} and ground) with only a small percentage of samples that fall close to the mid-supply point and hence potentially unstable at the output of the inverter due to time-varying variation sources (e.g., temperature, voltage, and noise). Fig. 5.4 illustrates the architecture of the 256-bit PUF module along with the readout circuitry.

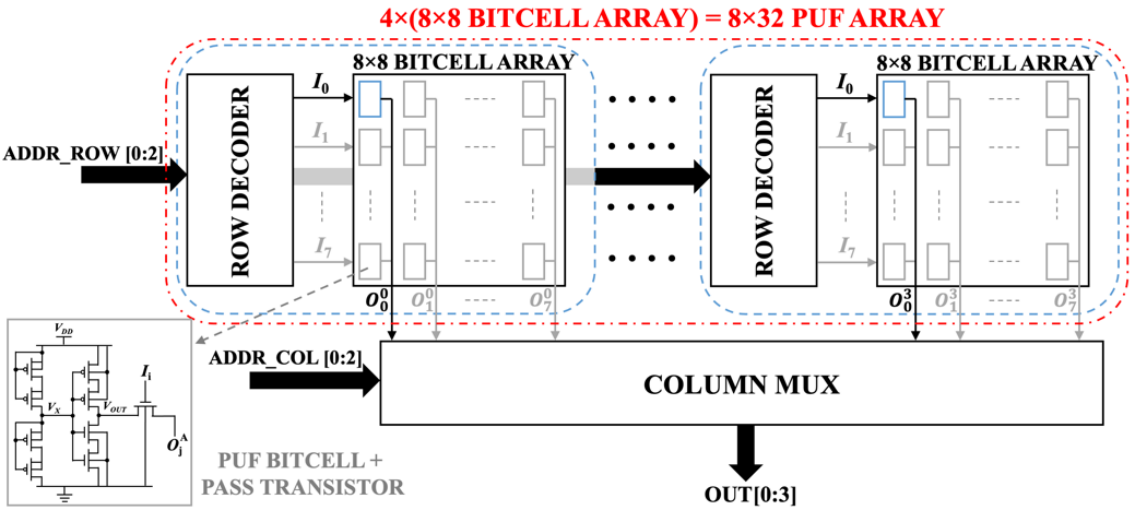


Fig. 5.4. Architecture of the PUF array.

From Fig. 5.4, The 256-bit array was organized in four 8 × 8 sub-blocks along with additional circuitry for readout. The bitcell consists of the circuit showed in Fig. 5.3(b), whose output node is connected to a pass-transistor for electrically isolating the bitcells belonging to the same column, as well as to select one row within the four 8 × 8 sub-blocks. In particular, the row selection is managed through a 3-to-8 input decoder according to the ADDR_ROW signal. Then an 8-to-1 output multiplexer carries out one output within the eight belonging to the same row for each sub-block according to the ADDR_COL signal. This results in an overall throughput of 4 bits per read. Fig. 5.5(a)-(d) report the measurements of the 8 × 32 PUF array at GK and different environmental conditions.

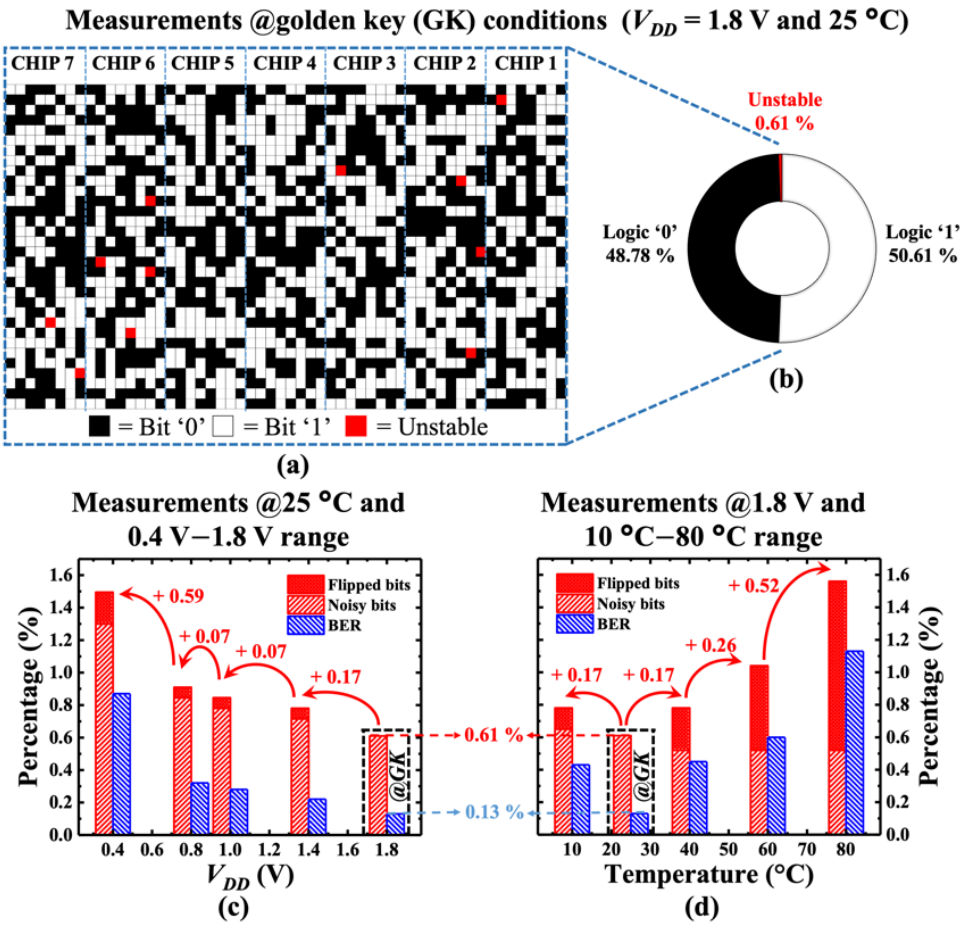


Fig. 5.5. Measurement of the 8×32 PUF array across seven test chips. (a) Logical speckle diagram and (b) breakdown among logic '1', logic '0', and unstable bits at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C). Percentage of unstable bits (i.e., flipped + noisy) under (c) voltage variations at T= 25 °C and (d) temperature variations at V_{DD}= 1.8 V.

Fig. 5.5(a) provides the logical speckle diagram of the seven test chips at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C). Fig. 5.5(b) report the breakdown between bit '1', bit '0', and unstable bits. The latter refers to the percentage of bits that flip at least once under 500 read evaluations. From this figure, at GK conditions the percentage of unstable bits is 0.61% which results to 0.13% of BER, thus proving the effectiveness of the proposed solution. Fig. 5.5(c) and (d) report the stability trend under voltage and temperature variations, respectively. From Fig. 5.5(c) the percentage of unstable bits (BER) increases from 0.61% (0.13%) up to 1.495% (0.87%) when decreasing the V_{DD} down to 0.4 V. From Fig. 5.5(d) the percentage of unstable bits increases when varying the temperature. In particular, the unstable bits increase from 0.61% (0.13%) up to 1.56% (1.13%) when increasing the temperature up to 80 °C. However, the overall instability under voltage variation is dominated by noisy bits while flipped bits increase up to 0.195% at V_{DD}= 0.4 V. On the other hand, the instability under temperature variations is dominated by flipped bits, especially under large variations. Indeed, the percentage of flipped bits increase up to 1.04% at T= 80 °C. Finally, Fig. 5.6(a)-(c) provide the measured PUF metrics at both GK and different environmental conditions across seven dice from 10,000 random CRPs. Fig. 5.6(a) reports the statistical distribution of the normalized inter-PUF HD at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C). From this figure the measured uniqueness of 0.493 is quite close to the ideal value of 0.5.

Fig. 5.6(b) provides the statistical distribution of the normalized intra-PUF HD at GK and different environmental conditions.

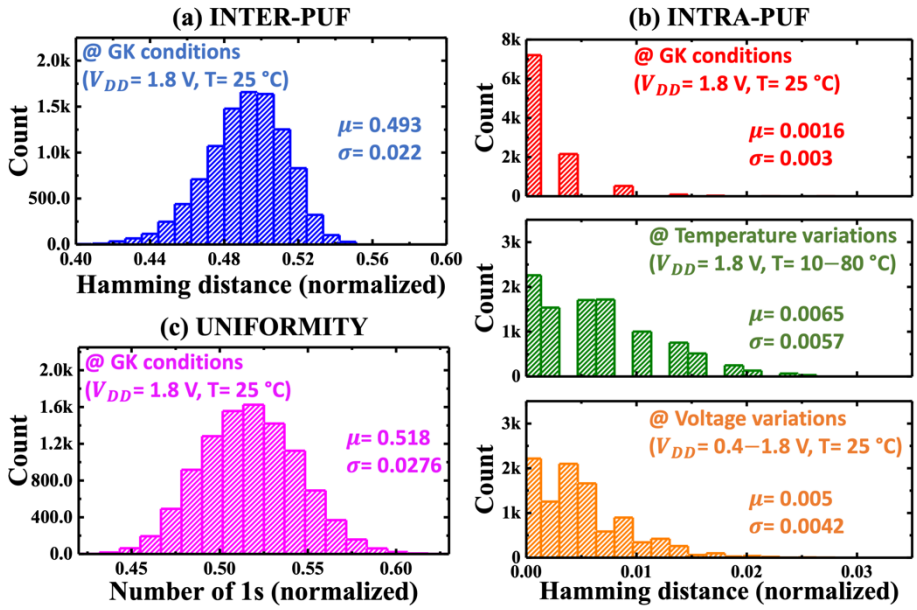


Fig. 5.6. Measured PUF metrics measured across seven dice from 10k random CRPs: (a) normalized inter-PUF HD at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C), (b) normalized intra-PUF HD at GK and different environmental (i.e., temperature and voltage) conditions, and (c) normalized number of bit '1' at GK conditions. Data were obtained using 32-bit PUF words.

From this figure the PUF instance at GK conditions the reproducibility of 0.9984 is quite close to the ideal value of 1, which leads to an identifiability (i.e., the ratio between inter and intra PUF) of 308×, thus proving the PUF ability of being distinguished from other PUF instances even under noisy conditions. From the same figure under voltage (temperature) variations the reproducibility of 0.995 (0.9935) remains very close to the ideal value of 1, thus leading to an identifiability of 99× (76×) and demonstrating the PUF ability of being distinguishable from the other instances even under voltage and temperature variations. Finally, Fig. 5.6(c) provides the probability of having bit '1' in the PUF response at GK conditions. From this figure the mean probability of 0.518, which is quite close to the ideal value of 0.5, at which corresponds a Shannon entropy of 0.9991 assesses the PUF uniformity. The latter is a fundamental metrics for ensuring an adequate degree of randomness. Moreover, randomness was more rigorously assessed by performing statistical NIST tests [25], which were passed by all measured dice.

5.3.2 Temporal majority voting module

Ideally, a PUF instance should have a reproducibility of 1, which means that the circuit should deliver the same response, for a given challenge, even under noisy or different environmental (i.e., voltage and temperature, VT) conditions. However, as shown above, the implemented PUF instance shows a small percentage of bits affected by on-chip noise VT variations thus indicating the need of using some stability enhancement technique. Anyway, the small percentage of noisy bits can be filtered out by using a temporal majority voting (TMV) module. This module collects q PUF responses, for a given challenge, and selects the response which outnumbers the others by

more than 50%. For example, considering q PUF words of 256-bit length of each bit the most frequent one will be chosen, i.e., the bit with an occurrence of $\frac{q}{2}$+1. It follows that the probability of choosing a distorted response is the BER raised to the *i-th* power, for all possible ways to choose an unordered subset of i responses from the set of q responses, and for all possible $i \in \left[\frac{q}{2}+1, q\right]$. This results in a probability P_q of choosing a stable PUF response for a given q, given by

$$P_q = 1 - \binom{q}{\frac{q}{2}+1} \cdot BER^{\frac{q}{2}+1} \quad (5.1)$$

From (5.1), for achieving a P_q close to 1, it is sufficient to have low q and low BER. Indeed, at GK conditions, for q values of 3 and 7 the values of P_q are 0.999995 and 0.9999999999, respectively, which are very close to the ideal value of 1.

5.3.3 Filter module

As described before, not all the 256-bit PUF responses are valid for implementing ECDSA-based cryptography. Indeed, both the private key ($P_r.K$) and the nonce (k) must fall in the range [1, n-1], where n indicates the order of elliptic curve. Indeed, both $P_r.K$ and k are generated from a truly random entropy source, but independently from each other, thus meaning that each 256-bit PUF response can assume values in the range of [0, 2^{256}-1]. Fig. 5.7 illustrates the architecture of the filter. The filter module receives a 256-bit string from the TMV module (i.e., the PUF response R) and sends it to the ECC module if it satisfies the ECC requirement (i.e., if it falls in the range between [1, n-1]). Otherwise, it starts the loop where different challenges will be sent to the PUF instance for generating different responses until one of them satisfies the ECC requirements. The same procedure is used for generating the nonce *k*.

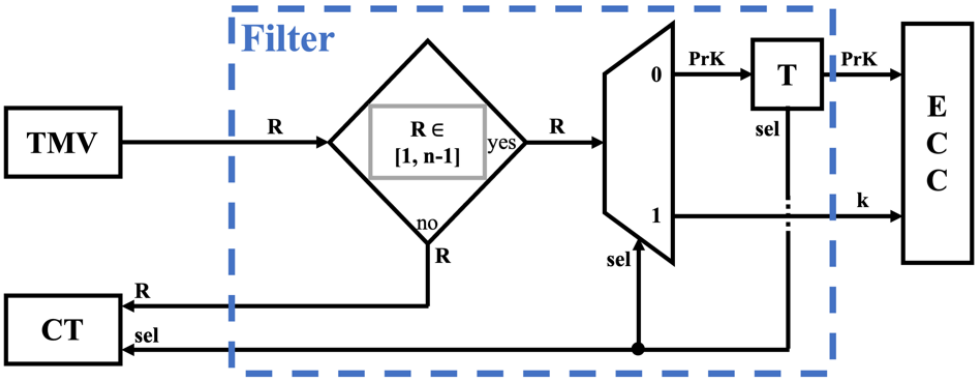


Fig. 5.7. The filter module architecture.

As shown in Fig. 5.7 a de-multiplexer switches between two different lines according to the required output (i.e., sel = 0 or 1 for sending the private key, PrK, or the nonce, k, respectively). The selector signal is determined by a toggle circuit (T) which nominally sets the selector signal to 0 except when the nonce is required. Moreover, the selector signal is also sent to the challenge trigger (CT) for challenging the PUF with the message m instead of that used for generating the

private key. It is important to stress that these generated numbers are deterministic, thus meaning that the same number of cycles is required for generating both the private key and nonce at each evaluation.

5.3.4 Challenge trigger module

The challenge trigger (CT) module controls the PUF input. Moreover, it sends at least two different challenges to the PUF module, one for generating the private key (i.e., PrK), when stimulated by one input message, and another function of the input message for generating the nonce (i.e., PrK). However, the CT module may submit further challenges if the current PUF response does not meet the ECC requirement described before. Fig. 5.8 illustrates the architecture of the challenge trigger (CT) module.

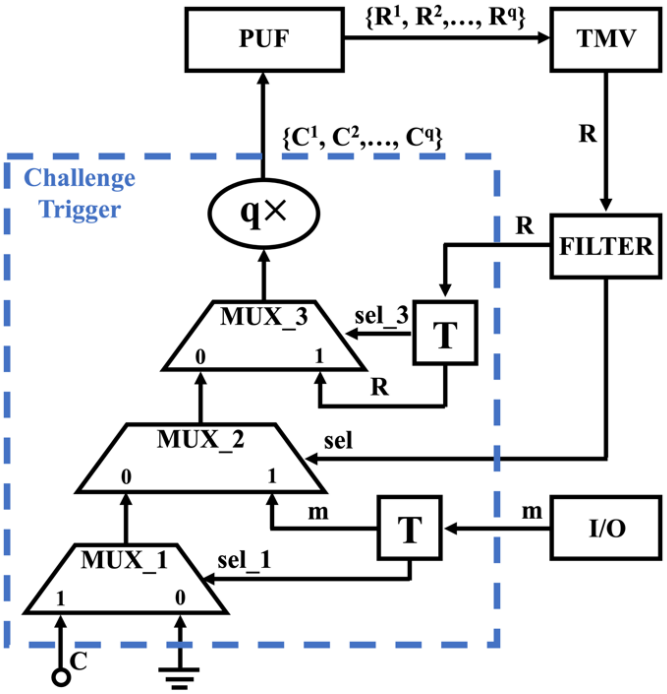


Fig. 5.8. Architecture of the challenge trigger module.

As shown in Fig. 5.8, the output of the CT is selected by a cascade of three 2-to-1 multiplexers. In particular, the first MUX (i.e., MUX_1) lets the hardwired challenge C to be fed into the PUF when sel_1 = 1, where this signal is set to '1' when the toggle circuit is stimulated by a message. The second MUX (i.e., MUX_2) chooses between the input challenge (i.e., when sel is equal to '0') and the message (i.e., when sel is equal to '1') where the selection signal is driven by the filter module upon the generation of the private key. Finally, the third MUX (i.e., MUX_3) chooses between the output of the second MUX (i.e., when sel_3 is equal to '0') and the "out-of-range" PUF response R (i.e., when sel_3 is equal to '1') when it does not satisfy the ECC requirement (i.e., of falling in the range of [1, n-1]). Once the filter identifies a proper challenge, the PUF evaluation is repeated q-times for enabling the temporal majority voting mechanism.

5.4 ECC component

As explained before, the ECC components includes the public key generator and the elliptic curve digital signature algorithm (ECDSA). The former generates the public key from the private one through the elliptic curve point multiplication (ECPM) for which $PuK = PrK \times G$, where G is the base point of the elliptic curve. On the other hand, the ECDSA generates the signed message (i.e., $(m)_{PrK}$ by using the input message (i.e., m), the private key (i.e., PrK), and a nonce (i.e., k). More specifically, the signed message is a pair of curve coordinates obtained by signing m with PrK. In this component the most expensive operation is the ECPM. Indeed, when implementing this operation in hardware is very important to consider some crucial points such as: (i) using as few registers as possible so that even if an attacker gets access to the tag he cannot extract sensitive information about the secret key; (ii) making the hardware solution resilience to side power attacks from which an attacker can extract sensitive information by observing the power consumption of the circuit. Our work was not focused on designing these components. However, one of the architectures proposed in [122]-[124] can be effectively used for implementing this part. Concerning the elliptic curve cryptography, among the infinite number of elliptic curves, only few of them are implemented in such algorithms. For example, Koblitz curves can be used for this architecture due to their fast and light computations which make them suitable for being implemented on smart devices.

5.5 I/O component

The I/O component has two main tasks. The first one is to enable the exchange of data with other objects (i.e., it receives a message as input and provides the signature of that message along with the public key). The second task is to provide the supply power to the passive tag. In particular, this identification tag is supposed to be intended for conferring identity to a physical or digital object. In the first case the I/O component enables a wireless communication while when the targeted object is digital, the tag is connected to the device through a wired communication bus. When considering a physical object, the identification procedure is ensured by dialoguing with an external reader through non-contact automatic identification technology such as RFID or NFC modules. In our case we adopted an NFC module since it is supported by several smartphones and IoT manufactures. In this way, when two devices which embed an NFC module are connected close to each other (i.e., around few centimeters), they establish a peer-to-peer connection which allows bidirectional communications.

5.6 Security analysis

This sub-chapter discusses the reliability of the system described above to the external attacks. More precisely, the attacks which can be performed during a prover-verifier interaction.

5.6.1 Memory leakage attacks

The first class belongs to memory leakage related attacks and refers to physical attacks. Since the tags are not tamper resistant, one attacker could physically access to the memory and extractive sensitive information by performing, for example, cold both attacks. This class of attacks is crucial when the secret key is stored in a NVM manner so that an attacker could extract secret data and performs actions like backward traceability (i.e., the knowledge of the internal state could help to identify past and future interactions) and cloning attacks (i.e., an attacker could clone the secret key and impersonate the tag itself). Anyway, the proposed tag architecture is resilience to these attack topologies, due to the use of PUF instead of conventional NVM-based approach, which ensures that the secret key is dynamically generated when required and not stored in NVMs.

5.6.2 Interception attacks

Interception attacks represent another class of possible attacks and refer to the interception of the communication messages between the prover and the verifier. This class of attacks could lead to actions like de-synchronization attacks (i.e., an attacker desynchronizes the update phase between the tag and server blocking the flow of messages) and man-in-the-middle attacks (i.e., an attacker put himself between the prover and verifier thus intercepting and/or modifying the exchanged messages). The latter could be particularly crucial when, during the communication phase, prover, and verifier exchange secret data with each other. However, in the adopted protocol only public information are sent during the authentication session.

5.6.3 Machine learning attacks

Another class topology is composed by machine learning attacks where an attacker collects a certain number of CRPs and tries to use them for building a mathematical model able to predict the PUF response for a given challenge. Anyway, the proposed tag architecture is inherently resilient to these attacks since the secret key is never outsourced. Moreover, the PUF itself cannot be mathematically predicted since each bit in the PUF array is generated independently from each other.

5.6.4 Replay attacks

Replay attacks refer to impersonating attacks where an attacker tries to impersonate the prover identity by reusing past messages obtained with eavesdropping attacks. This class of attacks can be counteracted by changing the message twice including a timestamp in the messages sent to the tags.

5.6.5 Spoofing attacks

Similar to the replay attacks, spoofing attacks lead to impersonate the tag identity. At first, an attacker can impersonate the verifier by sending the verifier by submitting a message to the tag

so that when the legitimate verifier queries the tag, the attacker can use the attacker sends the obtained response to the verifier. These attacks succeed only if the attacker sends the same message of the verifier to the tag. However, this happens only when the verifier generates messages by means of flawed PRNG which is not our case.

5.6.6 Insider attacks

Insider threats refer to attacks in which a malicious user breaks into the server and steal the information stored in it. However, the proposed tag is inherently robust to these attacks since no secret data are stored in the verifier side. Indeed, the public key without the private one is useless for the purpose of impersonating their real owners.

5.6.7 ECDSA attacks

Another class of attacks refers to that on the ECDSA algorithm. Moreover, this algorithm requires the private key, the message, and nonce. The nonce generation is crucial, since attackers would be able to infer the tag's private key in different way such as: (*i*) exploiting potential TRNG backdoors for inferring the nonce generation; (*ii*) collecting and comparing different $< m, (m)_{PrK} >$ pairs when they are generated using the same nonce; (*iii*) if the same *j* nonce are generated by *j* different tags, an attacker can simply solves a system of *j* linearly independent equations for inferring the private key. This attack topology is well-counteracted by the proposed tags since the use of PUF as TRNG element instead of a PRNG inherently ensures that different PUF instances show different behavior, thus indicating that the probability of generating the same nonce, for a given, challenge is close to zero. Moreover, since the PUF is challenged with the message for generating the nonce, its deterministic and unique behavior ensures that there is no possibility of having the same nonce with different messages.

5.6.8 Traceability attacks

Finally, Traceability attacks refer to the possibility of an attacker of tracing the tags locations, thus violating tags' owner privacy. However, our authentication protocol permits a tag to be authenticated by whoever capable of sending an input message to trigger its internal circuits and then obtain a response, which makes the tag prone to the traceability attack. The simplicity is achieved at the cost of lower privacy. Traceability can be avoided by means of a mutual authentication protocol (i.e., the tag can exchange private information upon the verifier authentication only). Unfortunately, a mutual authentication protocol requires tags to feature more complex characteristics, such as the use of a memory to store legitimate verifiers' data, and a control unit to execute a communication protocol. Our tag architecture, thus, must be employed where privacy does not represent an issue.

5.7 Conclusion

This chapter provides an application scenario for the proposed PUF cell. More precisely, the circuit described in Chapter 3 is employed for designing a smart tag architecture. It is composed by a set of digital components and designed for enabling reliable authentication task to be performed without the use of any memory and control unit. The authentication protocol is implemented by a set of cascade hardwired functions enabled by an external stimulus. As output a signed data as a proof of identity is delivered. The highly stable PUF, used for the generation of both private key and nonce, along with the memory less architecture and the implemented protocol, which does not store any secret data with/to the verifier, guaranty high reliability to external attacks during the prover-verifier interaction.

Chapter 6
Conclusion and future work

This thesis presents a novel class of static monostable PUF bitcell for hardware security applications. The adoption of a monostable circuit ensures that the output is always delivered even when the on-chip noise occasionally flips the output bit. Moreover, embedding the conversion block in the bitcell also ensures higher reliability and better uniqueness compared to solutions in which the comparator is shared within the column or the whole array. The proposed solution has been later implemented with emerging paper based MoS_2 FET along with other relevant PUF circuits for assessing the effectiveness of this emerging technology to be used for hardware security purposes. Finally, the proposed PUF architecture was employed for implementing a smart tag. This chapter summarizes the main achieved results also suggesting possible future developments.

6.1 Voltage divider based PUFs

In a more general point of view the proposed bitcell consists of a sub-threshold voltage divider between two nominally identical sub-circuits as core block along with an output inverter for the bit generation. The core block transforms the process variations into a voltage signal whereas the output inverter translates such voltage into a binary response. The adoption of a voltage divider between two nominally identical sub-circuits ensures an adequate degree of randomness regardless of the PVT conditions and good reliability to the VT variations, since they nominally have the same VT sensitivity. In this thesis, different circuital variants have been explored with the aim of improving the PUF stability while keeping the area degradation low.

The first explored solution consists of a voltage divider between two zero-V_{SG} pMOS where the mismatch between the two transistors is translated into a voltage deviation from the mid-supply point as far as low is the DIBL effect of the devices. Measurements of 20 samples demonstrated that the randomness is always ensured regardless of the VT conditions. However, due to both deep sub-threshold operations and mismatch in the voltage (e.g., DIBL effect) and temperature (e.g., V_{TH} temperature coefficient) sensitivities, the solution showed some unstable bits under VT variations. For this reason, different circuital variants were proposed. The first consists of including a negative-V_{SG} transistor for each sub-circuit with the aim improving the bitcell resiliency to the VT variations. Indeed, the addition of these transistors amplifies the transistor differences as well as shields such mismatch against the voltage variations. Measurements results across seven dice in 180-nm CMOS technology demonstrate the effectiveness of the proposed solution showing an overall native (i.e., without using stability enhancement techniques) instability at GK conditions (i.e., V_{DD}= 1.8 V and T= 25 °C) of 0.61 % with a BER of 0.13% which is strong competitive, compared to the prior art solutions. Moreover, the circuit showed a low instability sensitivity to VT variations. Indeed, the percentage of unstable bits (BER) increases up to 1.495% (0.87%) when decreasing the V_{DD} down to 0.4 V, and increases up to

1.56% (1.13%) when increasing the temperature up to 80 °C. Anyway, the instability under voltage variations is dominated by noisy bits (i.e., bits can flip at least once during different evaluations) which can be easily corrected by using majority voting techniques. On the other hand, the instability under temperature variations is dominated by flipping bits (i.e., bits that permanently flip under temperature variations) thus requiring more complex techniques for correcting reducing the KER. These results were achieved with a relatively large bitcell footprint compared to other solutions. Anyway, reducing the native instability is as important as improving the area-efficiency since the large amount of area and energy required by the error correction circuits. Area overhead can be improved by relaxing the inverter design as well as connecting the body effect of each transistor in the sub-circuit to each other. Simulation results demonstrates that around 38% of area saving can be achieved at the cost of a slightly increase of both overall instability and power consumption. On the other hand, native stability can be improved by considering more stacked solutions. More precisely, one or two negative-V_{SG} can be included in each sub-circuit for improving the overall gain of the circuit and the shielding effect on the M1-M2 mismatch against the voltage variations. Simulation results demonstrates the possibility of achieving a near-zero instability across a wide range of voltages at the cost of larger bitcell area.

6.2 PUF circuit implementation in 2D electronics

The proposed solution along with other relevant works were implemented with paper based MoS_2 FETs. More precisely, experimental measured I_D vs V_{DS} at different V_{GS} were employed for building a LUT-based Verilog-A model which was then imported into Cadence Virtuoso environment for enabling circuit simulations. The model also included information on the process variability extracted from the I_D vs V_{GS} at V_{DS}= V_{DD} of 27 nFETs from the same manufacturing lot. Simulation results demonstrate the possibility of using these emerging devices as building block for next-generation flexible electronics targeting hardware security applications.

6.3 PUF-based smart tag

The proposed PUF circuit was also involved for designing smart tag for anti-counterfeiting applications. The tag architecture includes a TRNG module for the generation of both private key and nonce (i.e., number used once) along with an ECC module for the public key generation and the digital signature of the message. The authentication protocol is implemented by a set of cascade hardwired functions enabled by an external stimulus. As output a signed data as a proof of identity is delivered. The highly stable PUF along with the proposed memory less architecture ensure high reliability to external software and hardware attacks.

6.4 Future work

The advent of the IoT scenario pushes the demand for having more and more secure systems. PUFs represent emerging cryptographic primitives whose scientific interest is bound to grow in

the next years. Nowadays, the proposed solutions show very interesting features along with several challenges. Indeed, the non-zero instability causes of using error correction or stability enhancement techniques which are not always feasible for IoT devices which operate with constrained cost and energy. As result, there are many possible future targets.

6.4.1 Technological level

The achieved results are also technology dependent. For this reason, testing and optimizing the proposed solutions in different technological nodes is one of the next targets. Moreover, in view of the growing scientific interest, another target is to design different PUF solutions with different technologies and with reconfigurable platforms, such as the FPGAs.

6.4.2 Circuit level

The proposed PUF solution show high native stability and low sensitivity to the VT variations. However, as described above, the overall voltage instability is dominated by noisy bits which could be easily corrected adopting a TMV. On the other hand, temperature instability is dominated by flipping bits which requires error correction schemes for reducing the BER. More stacked solution can be employed for suppressing the voltage sensitivity. However, they cannot be used for improving the stability under temperature variations. Since the zero-instability is required for making a PUF suitable for cryptographic applications one of the next targets is to suppress the temperature sensitivity. A possible solution could be, for example, making this cell configurable so that during the testing time we can choose the more stable M1-M2 pairs thus avoiding or at least reducing the overall instability under temperature variations.

6.4.3 Application level

At the application level, these circuits can be employed for different applications, as described in Chapter 2, such as cryptographic key generation, low-cost authentication, hardware assisted cryptographic protocols, etc. In Chapter 5 a smart tag for anti-counterfeiting applications was described. However, some applications are strictly related to the PUF performance. For example, the proposed PUF solution shows very low sensitivity to the voltage variations in terms of unstable bits, which means that errors induced by these variations can be easily corrected by using TMV techniques. On the other hand, the instability under temperature variations is dominated by flipping bits which cannot be corrected by using TMV approach. However, the relative high sensitivity to the temperature variations along with a low sensitivity to the voltage variations can be exploited for implementing a remote secure temperature sensor.

Bibliography

[1] K. Yang, D. Blauuw, and D. Silvester, "*Hardware Design for Security in Ultra-Low-Power IoT Systems: An Overview and Survey*," *IEEE Micro*, vol. 37, no. 6, pp. 72-89, 2017.

[2] W. Stallings, *"Cryptography and Network Security: Principles and Practice", Pearson, 2020.*

[3] I. Verbauwhede, *"Security adds an extra dimension to IC design: Future IC design must focus on security in addition to low power and energy", IEEE Solid-State Circuits Magazine,* vol. 9, no. 4, pp. 41–45, Nov. 2017.

[4] M. Rostami, F. Koushanfar, and R. Karri, "*A Premier on Hardware Security: Models, Methods, and Metrics"*, *Proceedings* of the IEEE, vol. 102, No. 8, pp. 1283-1295, 2014.

[5] S. Satpathy, S. Mathew, V. Suresh, M. Anders, H. Kaul, A. Agarawal, S. Hsu, G. K. Chen, R. Krishnamurthy, "*250mV-950mV 1.1 Tbps/W Double-Affine Mapped Sbox Based Composite-Field SMS4 Encrypt/Decrypt Accelerator in 14nm Tri-Gate CMOS*", in Proc. VLSI Circuits, 2016.

[6] M. M Islam, M. S. Hossain, M. Shahjalal, M. K. Hasan, and Y. M. Jang, "*Area-Time Efficient Hardware Implementation of Modular Multiplication for Elliptic Curve Cryptography*", IEEE Access, vol. 8, pp. 73898-73906, 2020.

[7] M. A. Mehrabi, C. Doche, and A. Jolfaei, "*Elliptic Curve Cryptography Point Multiplication Core for Hardware Security Module*", IEEE Transactions on Computer, vol. 69, no. 11, pp. 1707-1718, 2020.

[8] R. Kumar, V. Suresh, M. Kar, S. Satpathy, M. A. Anders, H. Kaul, A. Agarwal, S. Hsu, G. K. Chen, R. K. Krishnamurthy, V. De, and S. K. Mathew, "*A 4900-μm^2 839-Mb/s Side-Channel Attack-Resistant AES-128 in 14-nm CMOS With Heterogeneous Sboxes, Linear Masked MixColumns, and Dual-Rail Key Addition*", IEEE Journal of Solid-State Circuits, vol. 55, no. 4, pp. 945-955, 2020.

[9] R. Kumar, X. Liu, V. Suresh, H. K. Krishnamurthy, S. Satpathy, M. A. Anders, H. Kaul, K. Ravichandran, V. De, and S. K. Mathew, "*A Time-/Frequency-Domain Side-Channel Attack Resistant AES-128 and RSA-4K Crypto-Processor in 14-nm CMOS*", IEEE Journal of Solid-State Circuits, vol. 56, no. 4, pp. 1141-1151, 2021.

[10] H. L. Pham, T. H. Tran, T. D. Phan, V. T. D. Le, D. K. Lam, and Y. Nakashima, "*Double SHA-256 Hardware Architecture With Compact Message Expander for Bitcoin Mining*", IEEE Access, vol. 8, pp. 139634-139646, 2020.

[11] L.-Y. Yeh, P.-J. Chen, C.-C. Pai, and T.T. Liu, "*An Energy-Efficient Dual-Field Elliptic Curve Cryptography Processor for Internet of Things Applications*", IEEE Transactions on Circuits and Systems—II, vol. 67, no. 9, pp. 1614-1618, 2020.

[12] Z. Liu, H. Seo, A. Castiglione, K.-K. R. Choo, and H. Kim, "*Memory-Efficient Implementation of Elliptic Curve Cryptography for the Internet-of-Things*", IEEE Transactions on Dependable and Secure Computing, vol. 16, no. 3, pp. 521-529, 2019.

[13] M. Vatalaro, R. De Rose, M. Lanuzza, and F. Crupi, "*Static CMOS Physically Unclonable Functions Based on 4T Voltage Divider With 0.6%−1.5% Bit Instability at 0.4−1.8 V Operation in 180 nm," IEEE Journal of Solid-State Circuits,* vol. 57, no. 8, pp. 2509-2520, 2022.

[14] M. Vatalaro, R. De Rose, M. Lanuzza, and F. Crupi, "*Stability-Area Trade-off in Static CMOS PUF Based on 4T Subthreshold Voltage Divider*," *29th IEEE International Conference on Electronics Circuits and Systems* (*ICECS 2022*), 2022.

[15] S. Conti, L. Pimpolari, G. Calabrese, R. Worsley, S. Majee, D. K. Polyushkin, M. Paur, S. Pace, D. H. Keum, F. Fabbri, and G. Iannaccone, "*Low- voltage 2D materials-based printed field-effect transistors for integrated digital and analog electronics on paper*," *Nature Communications*, 11(1), 2020.

[16] M. Vatalaro, R. De Rose, M. Lanuzza, G. Iannaccone, and F. Crupi, "*Assessment of 2D-FET Based Digital and Analog Circuits on Paper,*" *Solid-State Electronics*, vol. 185, 2021.

[17] M. Vatalaro**,** R. De Rose, M. Lanuzza, P. Magnone, S. Conti, G. Iannaccone, and F. Crupi, "*Assessment of Paper Based MoS_2 FET for Physically Unclonable Functions*," *Solid-State Electronics*, vol. 194, 2022.

[18] M. Alioto, "*Enabling the Internet of Things-From Integrated Circuits to Integrated System,"* Cham, Switzerland: *Springer*, 2017.

[19] Y. Gao, S. F. Al-Sarawi and D. Abbott, "*Physical Unclonable Functions,*" *Nature Electronics*, vol. 3, pp. 81-91, 2020.

[20] M. Alioto, "*Aggressive Design Reuse for Ubiquitous Zero-Trust Edge Security -From Physical Design to Machine Learning-Based Hardware Patching*," in *IEEE Open Journal of the Solid-State Circuits Society*, 2022, doi: 10.1109/OJSSCS.2022.3223274.

[21] M. Alioto, "*Trends in Hardware Security from Basics to Asics*," *IEEE Solid-State Circuits Magazine*, vol. 11, no. 3, pp. 56-74, 2019.

[22] C. Herder, M. -D. Yu, F. Koushanfar and S. Devadas, "*Physical Unclonable Functions and Applications: A Tutorial*," in *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1126-1141, 2014.

[23] B. Halak, "*Physically Unclonable Functions from Basics Design Principles to Advanced Hardware Security Applications*," *Springer*, 2018.

[24] M. J. M. Pelgrom, A. C. J. Duinmaijer and A. P. G. Welbers, "*Matching properties of MOS transistors*," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1433-1439, 1989.

[25] A. Rukhin *et al.*, "*A statistical test suite for random and pseudorandom number generators for cryptographic applications*", National Institute of Standards and Technology, Gaithersburg, MD, USA, Tech. Rep. 800-22 Rev 1a, Sep. 2010, p. 131.

[26] G. E. Suh and S. Devadas, "*Physical Unclonable Functions for Device Authentication and Secret Key Generation*," *44th ACM/IEEE Design Automation Conference*, pp. 9-14, 2007.

[27] L. Lu, and T. T.-H. Kim, "*A programmable 6T SRAM-based PUF with dynamic stability data masking*," *IEEE Asian Solid-State Circuits Conference (A-SSCC)*, pp. 1-3, 2021.

[28] S. Taneja, V. K. Rajanna, and M. Alioto, "*In-memory unified TRNG and multi-bit PUF for ubiquitous hardware security*," *IEEE Journal of Solid-State Circuits, vol.* 57, no. 1, pp. 153-166, 2022.

[29] K. Liu, Y. Min, X. Yang, H. Sun, and H. Shinohara, "*A 373-F² 0.21%-Native-BER EE SRAM Physically Unclonable Function With 2-D Power-Gated Bit Cells and V_{SS} Bias-Based Dark-Bit Detection*," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 1719-1732, 2020.

[30] K. Liu, X. Chen, H. Pu, and H. Shinohara, "*A 0.5-V Hybrid SRAM Physically Unclonable Function Using Hot Carrier Injection Burn-In for Stability Reinforcement," IEEE Journal of Solid-State Circuits*, vol. 56, no. 7, pp. 2193-2204, 2021.

[31] J.-W. Nam, J.-H. Ahn, and J.-P. Hong, "*Compact SRAM-Based PUF chip employing body voltage control technique*." *IEEE Access*, vol. 10, pp. 22311-22319, 2022.

[32] Y. Shifman, A. Miller, Y. Weizman, A. Fish, and J. Shor, "*An SRAM PUF with 2 Independent Bits/Cell in 65nm*," *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1-5, 2019.

[33] L. Lu, and T. T. -H. Kim, "*A Sequence-Dependent Configurable PUF Based on 6T SRAM for Enhanced Challenge Response Space*," *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1-5, 2019.

[34] J. Li, T. Yang, M. Yang, P. R. Kinget, and M. Seok, "*An Area-Efficient Microprocessor-Based SoC With an Instruction-Cache Transformable to an Ambient Temperature Sensor and a Physically Unclonable Function*," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 3, pp. 728-737, 2018.

[35] Y. Shifman, A. Miller, O. Keren, Y. Weizman, and J. Shor, "*An SRAM-Based PUF With a Capacitive Digital Preselection for a 1E-9 Key Error Probability*," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 12, pp. 4855-4868, 2020.

[36] S. R. Sahoo, S. Kumar, and K. Mahapatra "*A Novel Configurable Ring Oscillator PUF With Improved Reliability Using Reduced Supply Voltage*," *Microprocessor and Microsystems*, vol. 60, pp. 40-52, 2018.

[37] J. Lee, D. Lee, Y. Lee, and Y. Lee, "*A 354F2 Leakage-Based Physically Unclonable Function With Lossless Stabilization Through Remapping for Low-Cost IoT Security*," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 2, pp. 648-657, 2021.

[38] C. Q. Liu, Y. Cao, and C. H. Chang, "*ACRO-PUF: A Low-power, Reliable and Aging-Resilient Current Starved Inverter-Based Ring Oscillator Physical Unclonable Function*," *IEEE Transactions on Circuits and Systems I*, vol. 64, no. 12, pp. 3138-3149, 2017.

[39] J. Zhang, X. Tan, Y. Zhang, W. Wang, and Z. Qin, "*Frequency Offset-Based Ring Oscillator Physical Unclonable Function*," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 4, no. 4, pp. 711-721, 2018.

[40] Z. -Y. Liang, H. -H. Wei, and T. -T. Liu, "*A Wide-Range Variation-Resilient Physically Unclonable Function in 28 nm*," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 3, pp. 817-825, 2020.

[41] G. Li, P. Wang, X. Ma, J. Lian, J. Shu, and Y. Zhang, "*A 215-F² Bistable Physically Unclonable Function With an ACF of <0.005 and a Native Bit Instability of 2.05% in 65-nm CMOS Process*," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 28, no. 11, pp. 2290-2299, 2020.

[42] S. Satpathy, S. Mathew, J. Li, P. Koeberl, M. Anders, H. Kaul, G. Chen, A. Agarwal, S. Hsu, R. Krishnamurthy, "*13fJ/bit probing-resilient 250K PUF array with soft darkbit masking for 1.94% bit-error in 22nm tri-gate CMOS*," *ESSCIRC 2014 - 40th European Solid State Circuits Conference (ESSCIRC)*, pp. 239-242, 2014.

[43] S. Satpathy, S. K. Mathew, V. Suresh, M. A. Anders, H. Kaul, A. Agarwal, S. K. Hsu, G. Chen, R. K. Krishnamurthy, and V. K. De, "*A 4-fJ/b Delay-Hardened Physically Unclonable Function Circuit With Selective Bit Destabilization in 14-nm Trigate CMOS*," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 4, pp. 940-949, 2017.

[44] S. K. Satpathy, S. K. Mathew, R. Kumar, V. Suresh, M. A. Anders, H. Kaul, A. Agarwal, S. Hsu, G. Chen, R. K. Krishnamurthy, and V. De, "*An All-Digital Unified Physically Unclonable Function and True Random Number Generator Featuring Self-Calibrating Hierarchical Von Neumann Extraction in 14-nm Tri-gate CMOS*," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 4, pp. 1074-1085, 2019.

[45] Y. Su, J. Holleman, and B. P. Otis, "*A Digital 1.6 pJ/bit Chip Identification Circuit Using Process Variations*," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 1, pp. 69-77, 2008.

[46] K. Yang, Q. Dong, D. Blaauw, and D. Sylvester, "*A 553F^2 2-Transistor Amplifier-Based Physically Unclonable Function (PUF) With 1.67% Native Instability*," *IEEE International Solid- State Circuits Conference (ISSCC 2017)*, pp. 146-148, 2017.

[47] D. Li, and K. Yang, "*A Self-Regulated and Reconfigurable CMOS Physically Unclonable Function Featuring Zero-Overhead Stabilization*," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 1, pp. 98-107, 2020.

[48] Y. He, D. Li, Z. Yu, and K. Yang, "*36.5 An Automatic Self-Checking and Healing Physically Unclonable Function (PUF) with <3×10-8 Bit Error Rate*," *IEEE International Solid- State Circuits Conference (ISSCC 2021)*, pp. 506-508, 2021.

[49] A. B. Alvarez, W. Zhao, and M. Alioto, "*Static Physically Unclonable Functions for Secure Chip Identification With 1.9–5.8% Native Bit Instability at 0.6–1 V and 15 fJ/bit in 65 nm*," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 3, pp. 763-775, 2016.

[50] S. Taneja, A. B. Alvarez, and M. Alioto, "*Fully Synthesizable PUF Featuring Hysteresis and Temperature Compensation for 3.2% Native BER and 1.02 fJ/b in 40 nm*," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 10, pp. 2828-2839, 2018.

[51] S. Taneja, and M. Alioto, "*PUF Architecture with Run-Time Adaptation for Resilient and Energy-Efficient Key Generation via Sensor Fusion*," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 7, pp. 2182-2192, 2021.

[52] P. Gan, X. Zhao, and Y. Cao, "*An All-MOSFET Voltage Reference-Based PUF Featuring Low BER Sensitivity to VT Variations and 163 fJ/Bit in 180-nm CMOS*," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 40, no. 6, pp. 1172-1182, 2021.

[53] G. Li, P. Wang, X. Ma, Y. Shi, B. Chen, and Y. Zhang, "*A Multimode Configurable Physically Unclonable Function With Bit-Instability-Screening and Power-Gating Strategies*," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 29, no. 1, pp. 100-111, 2021.

[54] B. Park, D. Forte, M. M. Tehranipoor, and N. Maghari, "*A Metal-Via Resistance Based Physically Unclonable Function With Backend Incremental ADC*," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 11, pp. 4700-4709, 202.

[55] X. Zhao, P. Gan, Q. Zhao, D. Liang, Y. Cao, X. Pan, and A. Bermak, "*A 124 fJ/Bit Cascode Current Mirror Array Based PUF With 1.50% Native Unstable Bit Ratio*," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 9, pp. 3494-3503, 2019.

[56] B. Karpinskyy, Y. Lee, Y. Choi, Y. Kim, M. Noh, and S. Lee, "*8.7 Physically unclonable function for secure key generation with a key error rate of 2E-38 in 45nm smart-card chips*," *IEEE International Solid-State Circuits Conference (ISSCC 2016)*, pp. 158-160, 2016.

[57] Y. Choi, B. Karpiniskyy, K.-M. Ahn, Y. Kim, S. Kwon, J. Park, Y. Lee, and M. Noh, "*Physically unclonable function in 28nm fdsoi technology achieving high reliability for aec-q 100 grade 1 and iso 26262 asil-b*," *IEEE International Solid- State Circuits Conference - (ISSCC)*, pp. 426-428, 2020.

[58] J. Li, and M. Seok, "*Ultra-Compact and Robust Physically Unclonable Function Based on Voltage-Compensated Proportional-to-Absolute-Temperature Voltage Generators*," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 9, pp. 2192-2202, 2016.

[59] J. Yoo, D. Kim, H. Park, M. Shim, C. Lee, and C. Kim, "*Physically Unclonable Function Using Ring Oscillator Collapse in 0.5 V Near-Threshold Voltage for Low-Power Internet of Things*," *IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia)*, pp. 206-212, 2018.

[60] J. Park, B. Kim, and J. -Y. Sim, "*A BER-Suppressed PUF With an Amplification of Process Mismatch Effect in an Oscillator Collapse Topology*," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 7, pp. 2208-2219, 2022.

[61] Y. Cao, L. Zhang, C. -H. Chang, and S. Chen, "*A Low-Power Hybrid RO PUF With Improved Thermal Stability for Lightweight Applications*," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 7, pp. 1143-1147, 2015.

[62] K.-H. Chuang, E. Bury, R. Degraeve, B. Kaczer, G. Groeseneken, I. Verbauwhede, and D. Linten, "*Physically unclonable function using CMOS breakdown position*,"*IEEE International Reliability Physics Symposium (IRPS 2017)*, pp. 4C-1.1-4C-1.7, 2017.

[63] K.-H. Chuang, E. Bury, R. Degraeve, B. Kaczer, T. Kallstenius, G. Groeseneken, D. Linten, and I. Verbauwhede, "*A multi-bit/cell PUF using analog breakdown positions in CMOS*," *IEEE International Reliability Physics Symposium (IRPS 2018)*, pp. P-CR.2-1-P-CR.2-5, 2018.

[64] K.-H. Chuang, E. Bury, R. Degraeve, B. Kaczer, D. Linten, and I. Verbauwhede, "*A Physically Unclonable Function Using Soft Oxide Breakdown Featuring 0% Native BER and 51.8 fJ/bit in 40-nm CMOS*," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 10, pp. 2765-2776, 2019.

[65] M. -Y. Wu, T.-H. Yang, L.-C. Chen, C.-C. Lin, H.-C. Hu, F.-Y. Su, C.-M. Wang, J. P.-H. Huang, H.-M. Chen, C. C.-H. Lu, E. C.-S. Yang, and R. S.-J. Shen, "*A PUF scheme using competing oxide rupture with bit error rate approaching zero*," *IEEE International Solid - State Circuits Conference - (ISSCC 2018)*, pp. 130-132, 2018.

[66] N. Liu, S. Hanson, D. Sylvester, and D. Blaauw, "*OxID: On-chip one-time random ID generation using oxide breakdown*," *Symposium on VLSI Circuits*, pp. 231-232, 2010.

[67] D. Jeon, J. H. Baek, Y.-D. Kim, J. Lee, D. K. Kim, and B.-D. Choi, "*A Physical Unclonable Function With Bit Error Rate* $< 2.3 \times 10^{-8}$ *Based on Contact Formation Probability Without Error Correction Code*," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 3, pp. 805-816, 2020.

[68] Z. He, W. Chen, L. Zhang, G. Chi, Q. Gao, and L. Harn, "*A Highly Reliable Arbiter PUF With Improved Uniqueness in FPGA Implementation Using Bit-Self-Test*," *IEEE Access*, vol. 8, pp. 181751-181762, 2020.

[69] Y. Hu, Y. Jiang, and W. Wang, "*Compact PUF Design With Systematic Biases Mitigation on Xilinx FPGAs*," *IEEE Access*, vol. 10, pp. 22288-22300, 2022.

[70] R. Della Sala, D. Bellizia, and G. Scotti, "*A Novel Ultra-Compact FPGA PUF: The DD-PUF*," *Cryptography*, 5, 23, 2021.

[71] Y. Cui, C. Gu, Q. Ma, Y. Fang, C. Wang, M. O'Neill, and W. Liu, "*Lightweight Modeling Attack-Resistant Multiplexer-Based Multi-PUF (MMPUF) Design on FPGA*," *Electronics*, 9, 815, 2020.

[72] R. Della Sala, D. Bellizia, and G. Scotti, "*A Lightweight FPGA Compatible Weak-PUF Primitive Based on XOR Gates*," *IEEE Transactions on Circuits and Systems−II*, vol. 69, no. 6, pp. 2972-2976, 2022.

[73] C. Gu, W. Liu, Y. Cui, N. Hanley, M. O'neill, and F. Lombardi, "*A Flip-Flop Based Arbiter Physically Unclonable Function (APUF) Design with High Entropy and Uniqueness for FPGA Implementation*," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 4, pp. 1853-1866, 2021.

[74] S. Zalivaka, A. A. Ivaniuk, and C.-H. Chang, "*Reliable and Modeling Attack Resistant Authentication of Arbiter PUF in FPGA Implementation With Trinary Quadruple Response*," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 4, pp. 1109-1123, 2019.

[75] M. A. Usmani, S. Keshavarz, E. Matthews, L. Shannon, R. Tessier, and D. E. Holcomb, "*Efficient PUF-Based Key Generation in FPGAs Using Per-Device Configuration*," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 27, no. 2, 2019.

[76] D. Divyanshu, R. Kumar, D. Khan, S. Amara, and Y. Massoud, "*Physically Unclonable Function Using GSHE Driven SOT Assisted p-MTJ for Next Generation Hardware Security Applications*," *IEEE Access*, vol. 10, pp. 93029-93038, 2022.

[77] R. A. John, N. Shah, S. K. Vishwanath, S. E. Ng, B. Febriansyah, M. Jagadeeswararao, C.-H. Chang, A. Basu, and N. Mathews, "*Halide Perovskite Memristor as Flexible and Reconfigurable Physical Unclonable Functions*," *Nature Communications*, 12:3681, 2021.

[78] A. Scholz, L. Zimmermann, U. Gengenbach, L. Koker, Z. Chen, H. Hahn, A. Sikora, M. B. Tahoori, and J. A.-Hagmann, "*Hybrid Low-Voltage Physical Unclonable Function Based on Inkjet-Printed Metal-Oxide Transistors*," *Nature Communications*, 11:5543, 2020.

[79] A. Dodda, S. S. Radhakrishnan, T. F. Schranghamer, D. Buzzell, P. Sengupta, and S. Das, "*Graphene-Based Physically Unclonable Functions That are Reconfigurable and Resilient to Machine Learning Attacks*," *Nature Electronics*, vol. 4, pp. 364-374, 2021.

[80] J. H. Kim, S. Jeon, J. H. In, S. Nam, H. M. Jin, K. H. Han, G. G. Yang, H. J. Choi, K. M. Kim, J. Shin, S.-W. Son, S. J. Kwon, B. H. Kim, and S. O. Kim, "*Nanoscale Physical Unclonable Function Labels Based on Block Co-Polymer Self-Assembly*," *Nature Electronics*, vol. 5, pp. 433-442, 2022.

[81] B. Gao, B. Lin, Y. Pang, F. Xu, Y. Lu, Y.-C. Chiu, Z. Liu, J. Tang, M.-F. Chang, H. Qian, H. Wu, "*Concealable Physically Unclonable Function Chip With a Memristor Array*," *Science Advances*, 8, eabn7753, 2022.

[82] H. M. Ibrahim, H. Abunahla, B. Mohammad, and H. AlKhzaimi, "*Memristor-Based PUF for Lightweight Cryptographic Randomness*," *Scientific Reports*, 12:8633, 2022.

[83] Q. Zhao, W. Zheng, X. Zhao, Y. Cao, F. Zhang, and M.-K. Law, "*A 108F²/Bit Fully Reconfigurable RRAM PUF Based on Truly Random Dynamic Entropy of Jitter Noise*," *IEEE Transactions on Circuits and Systems−I*, vol. 67, no. 11, 2020.

[84] B. Gao, B. Lin, X. Li, J. Tang, H. Qian, and H. Wu, "*A Unified PUF and TRNG Design Based on 40-nm RRAM With High Entropy and Robustness for IoT Security*," *IEEE Transactions on Electron Devices*, vol. 69, no. 2, pp. 536-542, 2022.

[85] Z. Cao, S. Zhang, J. Zhang, N. Xu, R. Li, Z. Guo, M. Song, Q. Zou, L. Xi, O. Lee, X. Yang, X. Zou, J. Hong, and L. You, "*Reconfigurable Physical Unclonable Function Based on Spin-Orbit Torque Induced Chiral Domain Wall Motion*," *IEEE Electron Device Letters*, vol. 42, no. 4, pp. 597-600, 2021.

[86] L. Zhang, X. Fong, C.-H. Chang, Z. H. Kong, and K. Roy, "*Highly Reliable Spin-Transfer Torque Magnetic RAM-Based Physical Unclonable Function With Multi-Response-Bits Per Cell*," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 8, pp. 1630-1642, 2015.

[87] A. T. Erozan, G. C. Marques, M. S. Golanbari, R. Bishnoi, S. Dehm, J. A.-Hagmann, and M. B. Tahoori, "*Inkjet-Printed EGFET-Based Physical Unclonable Function−Design, Evaluation, and Fabrication," IEEE Transactions on Very Large Scale Integrations (VLSI) systems*, vol. 26, no. 12, pp. 2935-2946, 2022.

[88] J. Das, K. S. Rajaram, D. Bugett, and S. Bhanja, "*MRAM PU: A Novel Geometry Based Magnetic PUF With Integrated CMOS*," *IEEE Transactions on Nanotechnology*, vol. 14, no. 3, pp. 436-443, 2015.

[89] L. Lin, S. Srivathsa, D. K. Krishnappa, P. Shabadi, and W. Burleson, "*Design and Validation of Arbiter-Based PUFs for Sub-45-nm Low-Power Security Applications*," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 4, pp. 1394-1403, 2012.

[90] K. Yang, Q. Dong, D. Blaauw, and D. Sylvester, "*14.2 A physically unclonable function with BER <10−8 for robust chip authentication using oscillator collapse in 40nm CMOS*," *IEEE International Solid-State Circuits Conference - (ISSCC 2015) Digest of Technical Papers*, pp. 1-3, 2015.

[91] Y. Cao, W. Zheng, X. Zhao, and C. -H. Chang, "*An Energy-Efficient Current-Starved Inverter Based Strong Physical Unclonable Function With Enhanced Temperature Stability*," *IEEE Access*, vol. 7, pp. 105287-105297, 2019.

[92] S. Jeloka, K. Yang, M. Orshansky, D. Sylvester and D. Blaauw, "*A sequence dependent challenge-response PUF using 28nm SRAM 6T bit cell*," *Symposium on VLSI Circuits*, pp. C270-C271, 2017.

[93] L. Lu, T. Yoo, and T. T. -H. Kim, "*A 6T SRAM Based Two-Dimensional Configurable Challenge-Response PUF for Portable Devices*," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 69, no. 6, pp. 2542-2552, 2022.

[94] Y. Cao, C. Q. Liu, and C. H. Chang, "*A Low Power Diode-Clamped Inverter-Based Strong Physical Unclonable Function for Robust and Lightweight Authentication*," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 11, pp. 3864-3873, 2018.

[95] Y. -C. Lai, C. -Y. Yao, S. -H. Yang, Y. -W. Wu, and T. -T. Liu, "*A Robust Area-Efficient Physically Unclonable Function With High Machine Learning Attack Resilience in 28-nm CMOS*," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 69, no. 1, pp. 347-355, 2022.

[96] H. Zhuang, X. Xi, N. Sun, and M. Orshansky, "*A Strong Subthreshold Current Array PUF Resilient to Machine Learning Attacks*," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 1, pp. 135-144, 2020.

[97] R. De Rose, F. Crupi, M. Lanuzza, and D. Albano, "*A Physical Unclonable Function Based on a 2-Transistor Subthreshold Voltage Divider*," *International Journal of Circuit Theory and Applications*, 45, pp. 260-273, 2017.

[98] C. Böhm, and M. Hofer, "*Physical Unclonable Functions in Theory and Practice*," New York, NY, USA: Springer, 2013.

[99] D. Akinwande, N. Petrone, and J. Hone, "*Two-Dimensional Flexible Nanoelectronics*," *Nature Communications*, 5:5678, 2014.

[100] A. D. Franklin, "*Nanomaterials in Transistors: From High-Performance to Thin-Film Applications*," *Science*, 329(6249), 2015.

[101] M. Chhowalla, D. Jena, and H. Zhang, "*Two-Dimensional Semiconductors for Transistors*," *Nature Reviews Materials*, 1:16052, 2016.

[102] G. Iannaccone, F. Bonaccorso, L. Colombo, and G. Fiori, "*Quantum Engineering of Transistors Based on 2D Materials.*

[103] E. G. Marin, D. Marian, M. Perucchini, G. Fiori, and G. Iannaccone, "*Lateral Heterostructures Field-Effect Transistors Based on Two-Dimensional Material Stacks With Varying Thickness and Energy Filtering Source*," *ACS Nano*, 14(2):1982-9, 2020.

[104] D. K. Polyushkin, S. Wachter , L. Mennel, M. Paur, M. Paliy, G. Iannaccone, G. Fiori, D. Neumaier, B. Canto, and T. Mueller, "*Analogue two-dimensional semiconductor electronics*," *Nature Electronics*, 3(8): 486–91, 2020.

[105] G. Iannaccone, Q. Zhang, S. Bruzzone, and G. Fiori, "*Insights on the physics and application of off-plane quantum transport through graphene and 2D materials*," *Solid-State Electronics*,115:213–8, 2016.

[106] S. Manzeli, D. Ovchinnikov, D. Pasquier, O. V. Yazyev, and A. Kis, "*2D transition metal dichalcogenides*," *Nature Reviews Materials*, 2:17033, 2017.

[107] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, and A. Kis, "*Single-layer MoS2 transistors*," *Nature Nanotechnology*, 6:147–50, 2011.

[108] D. Marian, E. Dib, T. Cusati, A. Fortunelli, G. Iannaccone, and G. Fiori, "*Two-dimensional transistors based on MoS2 lateral heterostructures*," *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016.

[109] O. Samy, S. Zeng, M. D. Birowosuto, and A. El Moutaouakil, "*A review on MoS2 properties, synthesis, sensing applications and challenges*," *MDPI Crystals*, 11(4):355, 2021.

[110] S. Park, and D. Akinwande, "*First demonstration of high performance 2D monolayer transistors on paper substrates*," *IEEE International Electron Devices Meeting (IEDM)*, 2017.

[111] S. Conti, L. Pimpolari, G. Calabrese, R. Worsley, S. Majee, D. K. Polyushkin, M. Paur, S. Pace, D. H. Keum, F. Fabbri, and G. Iannaccone, "*Low- voltage 2D materials-based printed field-effect transistors for integrated digital and analog electronics on paper*," *Nature Communications*, 11(1), 2020.

[112] D. Tobjo¨rk, and R. O¨sterbacka, "*Paper Electronics*," *Advanced Materials*, 23(17):1935–61, 2011.

[113] Y. Zhang, L. Zhang, K. Cui, S. Ge, X. Cheng, M. Yan, J. Yu, and H. Liu, "*Flexible Electronics Based on Micro/Nano Structured Paper*," *Advanced Materials*, 30(51):1801588, 2018.

[114] S. Huang, Y. Liu, Y. Zhao, Z. Ren, and C. F. Guo, "Flexible Electronics: Stretchable Electrodes and Their Future," Advance Functional Materials, 29(6):1805924, 2019.

[115] A. Ortiz-Conde, F. J. García-Sa´nchez, J. Muci, A. Ter´an Barrios, J. J. Liou, and C.-S. Ho, "*Revisiting MOSFET threshold voltage extraction methods*,". *Microelectron Reliability*, 53(1):90–104, 2013.

[116] S. Strangio, P. Palestri, M. Lanuzza, D. Esseni, F. Crupi, and L. Selmi, "*Benchmarks of a III-V TFET technology platform against the 10-nm CMOS FinFET technology node considering basic arithmetic circuits*," *Solid-State Electronics*, 128:37–42, 2017.

[117] S. Strangio, F. Settino, P. Palestri, M. Lanuzza, F. Crupi, D. Esseni, and L. Selmi, "*Digital and analog TFET circuits: design and benchmark*," *Solid-State Electronics*, 146:50–65, 2018.

[118] A. Falcone, C. Felicetti, A. Garro, A. Rullo, and D. Sacca`, "*PUF- based Smart Tags for Supply Chain Management*," in *the 16th International Conference on Availability, Reliability and Security*, 2021, pp. 1–7.

[119] D. Vujičić, D. Jagodić, and S. Ranđić, "*Blockchain technology, bitcoin, and Ethereum: A brief overview*," *17th IEEE international symposium infoteh-jahorina (infoteh),* 1–6, 2018.

[120] N. Alzahrani, and N. Bulusu, "*Block-supply chain: A new anti- counterfeiting supply chain using NFC and blockchain*," *Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems*, 30–35, 2018.

[121] K. Toyoda, P. T. Mathiopoulos, I. Sasase, and T. Ohtsuki, "*A novel blockchain-based product ownership management system (POMS) for anti-counterfeits in the post supply chain*," *IEEE access*, vol. 5, pp. 17465–17477, 2017.

[122] M. S. Hossain, E. Saeedi, and Y. Kong, "Parallel Point- multiplication Architecture Using Combined Group Operations for High-speed Cryptographic Applications," *Plos one*, vol. 12, no. 5, p. e0176214, 2017.

[123] N. Koblitz, "CM-curves with Good Cryptographic Properties," in *Annual international cryptology conference*. Springer, 1991, pp. 279– 287.

[124] L.-Y. Yeh, P.-J. Chen, C.-C. Pai, and T.-T. Liu, “An Energy-efficient Dual-field Elliptic Curve Cryptography Processor for Internet of Things Applications,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 9, pp. 1614–1618, 2020.

List of Publications

International Conferences

1) C. Felicetti, A. Furfaro, D. Saccà, **M. Vatalaro**, M. Lanuzza, F. Crupi, "*Making IoT Services Accountable: a Solution Based on Blockchain and Physically Unclonable Functions*", *12th International Conference on Internet and Distributed Computing Systems* (*IDCS 2019*), Napoli, Italy, 2019.

2) **M. Vatalaro**, R. De Rose, M. Lanuzza, F. Crupi*, "Stability-Area Trade-off in Static CMOS PUF Based on 4T Subthreshold Voltage Divider",* Accepted on the *29th IEEE International Conference on Electronics Circuits and Systems* (*ICECS 2022*)*,* Glasgow*,* UK, 2022*.*

International Journals

1) **M. Vatalaro**, T. Moposita, S. Strangio, L. Trojman, A. Vladimirescu, M. Lanuzza, F. Crupi, "*A Low-Voltage, Low-Power Reconfigurable Current-Mode Softmax Circuit for Analog Neural Networks*" *Electronics*, vol. 10, no. 9, p. 1004, 2021.

2) **M. Vatalaro**, R. De Rose, M. Lanuzza, G. Iannaccone, F. Crupi, "*Assessment of 2D-FET Based Digital and Analog Circuits on Paper*", *Solid-State Electronics*, vol. 185, 2021.

3) **M. Vatalaro**, R. De Rose, M. Lanuzza, F. Crupi, "*Static CMOS Physically Unclonable Function Based on 4T Voltage Divider With 0.6%-1.5% Bit Instability at 0.4-1.8 V Operation in 180 nm*" *IEEE Journal of Solid-State Circuits*, vol. 57, no. 8, pp. 2509-2520, 2022.

4) **M. Vatalaro**, R. De Rose, M. Lanuzza, P. Magnone, S. Conti, G. Iannaccone, F. Crupi, "*Assessment of Paper-Based MoS_2 FET for Physically Unclonable Functions*", *Solid-State Electronics*, vol. 194, 2022.

5) A. Rullo, C. Felicetti, **M. Vatalaro,** R. De Rose, M. Lanuzza, F. Crupi, and D. Saccà, "*PUF-Based Authentication-Oriented Architecture for Identification Tags," Submitted to IEEE Transactions on Dependable and Secure Computing*.