Reliability vs. Security: Challenges and Opportunities for Developing Reliable and Secure Integrated Circuits

(Invited Paper)

Fahim Rahman, Domenic Forte, and Mark M Tehranipoor
Department of Electrical and Computer Engineering, University of Florida
Email: fahim034@ufl.edu, {dforte, tehranipoor}@ece.ufl.edu

***Abstract*—As technology further scales, devices offer better performance with faster speed and lower power albeit at the cost of reliability. Advanced technology nodes introduce higher variations in manufacturing processes, and devices experience greater aging and environmental degradation. Although such reliability issues should be suppressed for the sake of performance in both CMOS and post-CMOS devices, researchers have leveraged them for a variety of applications and unique primitives for hardware-oriented security. In this paper, we present a comprehensive study on device reliability and security, and make a qualitative assessment of different variability and degradation sources based on their impact on performance, reliability and security. We conclude that reliability and security both play vital roles for respective applications and must be treated in a holistic manner. Hence we urge the reliability and security communities to work together to develop new technologies for designing high performance, reliable and secure integrated circuits.**

***Index Terms*—Hardware Security, Reliability, Process Variation, Aging, Physical Unclonable Functions, Counterfeit Electronics, Hardware Trojan.**

I. Introduction

In the advancement of CMOS technology, integrated circuits (ICs) have continued to achieve better performance at faster speed and lower power with device feature size shrinking through each technology node. However, in advanced nodes with smaller feature sizes, the industry also faces major manufacturing and reliability issues [1], [2]. In particular, CMOS devices are now experiencing greater process variations, and more aggressive degradation due to aging and performance variation with temperature and power supply fluctuations. Such a degradation can have a drastic negative effect on IC performance and reliability, and proper solutions are still being sought [3].

Meanwhile, with the ubiquity of electronics in our daily lives, there is an increased demand for secure and reliable system design. In particular, the area of hardware-oriented security – by offering new primitives and numerous countermeasures to withstand emerging vulnerabilities and attacks – has seen tremendous growth over the past decade [4]. Today most of the hardware security applications and primitives heavily rely on CMOS platforms, and more interestingly, many of them seeks to leverage the variability and reliability phenomena that designers often suppress for the sake of performance. For example, hardware security primitives such as physical unclonable functions (PUFs) and true random number generators (TRNGs) leverage inherent process variation of the device to extract entropy [5]–[7]. Aging and wear-out mechanisms, that pose undesirable degradation in terms of device performance, can be exploited to detect counterfeit electronics which is a major threat to electronic component supply chain security [8]. Further with the rise of emerging threats and vulnerabilities such as hardware Trojans [9], and longstanding attacks becoming more practical, we constantly seek for new primitives and countermeasures, and look for the opportunities that come from the device's inherent properties to enhance security. Researchers are also investigating emerging post-CMOS nanoscale devices, such as phase change memory and carbon nanotubes, since they offer interesting properties for hardware-oriented security. Hence we find that reliability issues play a crucial role in developing hardware security primitives and applications, whereas such issues might not be as desirable for conventional high performance applications.

In this paper, we present a comprehensive study on device reliability and security and their crossroad, and shed light on the challenges and opportunities they offer. For this, we make a qualitative assessment of different variability and degradation sources based on their impact on performance, reliability and security. We find that no one particular mechanism works the best for all, and performance, reliability and security must be treated in a holistic way. Therefore we urge the reliability and security communities to work together to develop new technologies for designing high performance, reliable and secure integrated circuits.

The rest of the paper is organized as follows. Section II provides preliminaries on some commonly discussed hardware security primitives, attacks, and countermeasures. Section III discusses reliability issues in modern technology nodes, and presents available opportunities and related challenges regarding hardware security and reliability. Section IV discusses PUF reliability, and offers possible solutions. Section V presents leveraging aging degradation for counterfeit electronics detection. Section VI discusses design and detection of reliability-based hardware Trojans. Section VII explores opportunities and challenges in emerging nano-devices in terms of reliability and security. Finally, Section VIII concludes our work.

II. Hardware Security Preliminaries

Researchers have proposed several hardware security primitives and countermeasures that offer safeguard to various threats and vulnerabilities. In this section, we briefly discuss some security primitives, attacks, and countermeasures.

This project was supported in part by an AFOSR MURI grant under award number FA9550-14-1-0351.

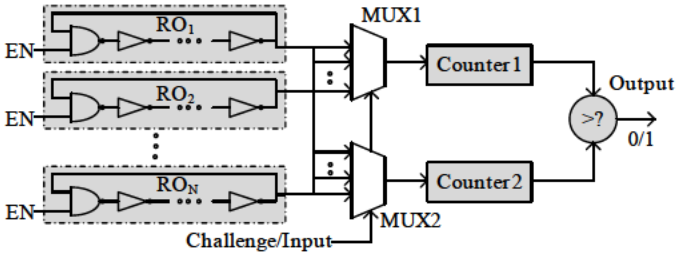

Fig. 1: Conventional RO-PUF [11].

A. Physical Unclonable Function

A Physical Unclonable Function (PUF) is a cryptographically secure one-way function that generates a digital output (response) for a given digital input (challenge) without revealing any predictable mapping between challenge and response [5], [6]. It has notable security-oriented applications such as key generation, and IC identification and authentication [10]. As the name suggests, a PUF's behavior is derived from inherent non-deterministic variations coming from the IC manufacturing process. Therefore circuits acting as PUFs can generate different challenge-response pairs (CRPs) although they are identical by design, and the CRPs cannot be predicted or cloned with another model or physical design. Figure 1 depicts a conventional Ring Oscillator PUF (RO-PUF), a widely popular architecture, that has N-identical ring oscillators each running with slightly different frequencies due to process variation [11]. Here, a challenge selects a pair of ROs, via multiplexers, whose frequencies are measured and compared to generate a response bit. Also PUF can produce responses on the fly, and hence offers an attractive volatile and tamper-resistant alternative to conventional non-volatile memory based key-storage approaches that are proven to be subject to tampering and imaging attacks [12].

B. True Random Number Generator

True random number generators (TRNGs) are well known for their wide use in cryptography and hardware security, most notably for generating session keys, random seeds, nonces, etc [15]. A TRNG leverages inherent entropy sources such as thermal noise, shot noise, flicker noise, random telegraph noise (RTN), power supply noise, clock jitter, etc. and translates such random physical phenomena into random bits [7], [13]. In contrast to pseudo-random number generators (PRNGs), which generates a seemingly random but deterministic bit-sequence based on a given seed, a TRNG should always show non-deterministic behavior. Such true randomness can be checked using statistical tests (e.g. NIST Test Suite [14]). In addition, a TRNG should have high throughput with minimal post-processing. However, TRNGs may show a certain level predictability, especially under environmental variations, which makes it vulnerable to various hardware-based attacks, such as power supply variation or frequency injection attacks, allowing the attacker to exploit the TRNG outcome [15], [16].

C. Counterfeit Electronics

Global electronic supply chain experiences different types of counterfeits, such as recycled, remarked, cloned, overproduced, and out-of-spec/defective ICs. Recycled and remarked ICs contribute to most of the counterfeit ICs present in the supply chain today. Counterfeit electronics pose a significant challenge due to the lack of efficient, robust, and low-cost detection and avoidance techniques. However, developing a comprehensive solution for different ICs (ranging from microprocessors to analog ICs) from different vendors is a challenging task. Moreover, this issue becomes more serious for legacy parts used in critical applications, and commercial-off-the-shelf (COTS) products that do not have proper documentation and/or supply chain history. Physical inspection and imaging based counterfeit detection has been quite successful till date, since it can detect the physical aspects of aging and wear-out mechanisms, such as bond wire defects, die cracks, etc., as well as clone, defective, and out-of-spec components. However, such techniques are extremely costly, slow, often destructive, and not suitable for large scale automated detection [17], [18]. Hence, investigation of electrical and parametric tests, and embedded anti-counterfeit designs are growing in popularity as fast and low-cost alternatives for detection and prevention of counterfeit electronics. However, choice of particular electrical characteristic and/or design is crucial for such applications since it may get affected by large process and environmental variations [19].

D. Hardware Trojans

With the globalization of IC design and manufacturing processes, there has been new paradigm in IC security and trust issues; especially when one chip now consists of a few to tens of IPs coming from different vendors and many parties are distributed across the globe. Further, the design house may not have access to a trusted foundry. Thus hardware Trojan has become one of the key challenges to hardware security and trust community. A hardware Trojan, in its simplest form, can be viewed as an unauthorized circuitry that is inserted into the original chip by adversaries without designer's knowledge to trigger specific actions. Such a Trojan may be tiny in size, can be inserted at any level of design abstraction (i.e. RTL, gate-level or layout), and most importantly can have virtually any analog or digital functionality [9]. Detecting such Trojans in a fabricated chip is extremely difficult, since the designer does not have any knowledge of it (e.g. whether the chip is actually Trojan-free or not; size, type, and location of the Trojan if inserted; behavior of the inserted Trojan; etc.). Hence finding new Trojan categories and detection and prevention of such Trojans have been some of the topmost concerns in the community.

III. Reliability and Security: Finding a Sweet Spot

Performance of conventional logic/memory applications mostly varies due to manufacturing process variation, run-time conditions and aging. In this section, we review some of these phenomena, and build a general picture of how these issues effect aforementioned security applications and primitives.

A. Process Variation

IC manufacturing process has numerous sources of systematic and random variations that play a critical role for yield and performance in semiconductor devices. CMOS front end of the line (FEOL) variation sources are most notably patterning (proximity) effects, line-edge and line-width roughness (LER, LWR), and gate dielectric such as oxide thickness variations,

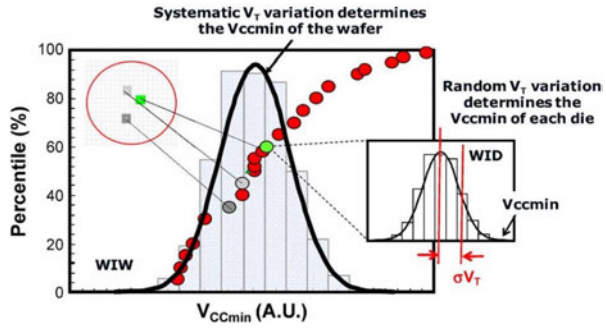

Fig. 2: SRAM minimum operating voltage (V_{CCmin}) is affected by systematic and random transistor variations [1], [21].

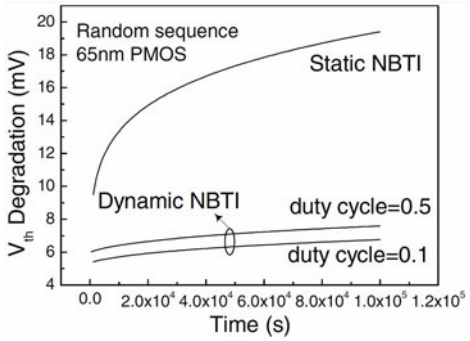

Fig. 3: V_{th} degradation for static and dynamic NBTI [26].

defects and traps. Variations due to random dopant fluctuations and gate material granularity (poly-si or metal gate), are also becoming significant in advanced nodes [1], [2]. Back end of the line (BEOL) sources, such as metal interconnect and dielectric variations, also have significant impact [20]. All these variations cause deviation in device characteristics making every single transistor slightly different from each other with a shift from the nominal performance. For example, Figure 2 depicts how SRAM cells experience a deviation in minimum operating voltage due to such process variations [1], [21]. Since such manufacturing process variations are undesirable, yet unavoidable, to achieve peak device and circuit performance, best efforts are required to model the these variation sources and related impacts, and suppress them to a small and acceptable range.

B. Environmental Variations

Similar to manufacturing process variation, run-time environmental variations, such as temperature and power supply noise, also have direct impact on the transistor electrical characteristics. These effects are more prominent in high performance and high speed designs in the advanced technology nodes. The operating temperature affects carrier mobility (μ) and threshold voltage (V_{th}) of the transistor, and thus impacts the speed (delay) of the device. An increase in the temperature decreases V_{th} which leads to an increase in the drain saturation current (I_{DS}) and leakage current. On the other hand, it decreases μ, which decreases I_{DS}. These two competing mechanisms eventually determine the speed variation of the transistor. However, technology nodes also play a crucial role to determine which of the two dominates. For example, technology nodes from 45-nm and below show an increase in device speed with temperature, whereas it is opposite in older technologies. In addition, global and local power supply noise have adverse impact on performance, since such variations also cause a shift in V_{th} and I_{DS} from the nominal value [22]. Hence, for both cases, a system is less robust. However, unlike a permanent shift in performance resulting from the manufacturing process variation or aging, a variation in the environmental condition causes a temporary change, and thus it is compensated once the device returns to its nominal operating state.

C. Aging and Wear-out Mechanisms

Aging degradation and wear-out mechanisms, such as bias temperature instability (BTI), hot carrier injection (HCI), electromigration (EM) and time dependent dielectric breakdown (TDDB), lead to poorer device and circuit performance. The magnitude of such degradation largely depends on the device workload, active bias and inherent random defects, and varies in different technology nodes [23].

1) Bias Temperature Instability: BTI is a key aging mechanism that slows down transistors by increasing V_{th} over time. Negative-BTI (NBTI) occurs in PMOS transistors for bias condition $V_{gs} = -V_{dd}$, known as stress phase, whereas positive-BTI (PBTI) occurs in NMOS transistors for the condition $V_{gs} = V_{dd}$. During the stress phase (t_{stress}), interface traps are generated at $SiO_2 - Si$ interface resulting in an increase in V_{th}. This is partially recoverable since some of the traps are recovered when $V_{gs} = 0$ at relaxation or recovery phase (t_{rec}). Typically NBTI is dominant compared to PBTI beyond $65nm$ technology nodes, however the latter it is eventually getting prominent for high-k metal-gate devices [24]. Increase in V_{th} induced by NBTI for a given stress is expressed as [25]:

$$\Delta V_{th_{stress}} = A_{NBTI} \times t_{ox} \times \sqrt{C_{ox}(V_{dd} - V_{th})} \times e^{\left(\frac{V_{dd}-V_{th}}{t_{ox}E_0} - \frac{E_a}{kT}\right)} \times t_{stress}^{0.25} \quad (1)$$

where t_{ox} is oxide thickness and C_{ox} is gate capacitance per unit area. E_0 and E_a are device dependent constants, A_{NBTI} is an aging rate dependent constant and k is Boltzmann constant. Overall degradation of ΔV_{th} considering stress and recovery is given by [25]:

$$\Delta V_{th} = \Delta V_{th_{stress}} \times (1 - \sqrt{\eta \times t_{rec}/(t_{stress} + t_{rec})}) \quad (2)$$

where η is device dependent recovery constant.

Since transistors experience NBTI in static and dynamic mode, overall degradation varies depending on the workload and operating condition, as shown in Figure 3 [26].

2) Hot Carrier Injection: HCI refers to the collision and injection of high energy carriers into $SiO_2 - Si$ interface at channel region due to high electric field. It generates charged defects in gate oxide and interface, hence the transistor experiences an increase in V_{th}. In addition, HCI reduces the mobility of a device, and decreases the on current (I_{ON}) [27]. V_{th} shift due to HCI can be empirically expressed as [25]:

$$\Delta V_{th_{HCI}} = A_{HCI} \times \alpha_s \times f \times t^{0.5} \times e^{\frac{V_{dd}-V_{th}}{t_{ox}E_l}} \quad (3)$$

where A_{HCI} is a technology dependent constant, α is activity factor, f is device operating frequency, t is total run-time and rest of the parameters hold the meaning as previously mentioned for NBTI. HCI is more prominent in NMOS with

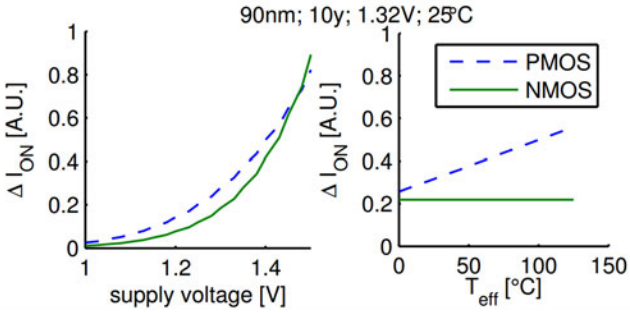

Fig. 4: Voltage and temperature dependence of HCI [27].

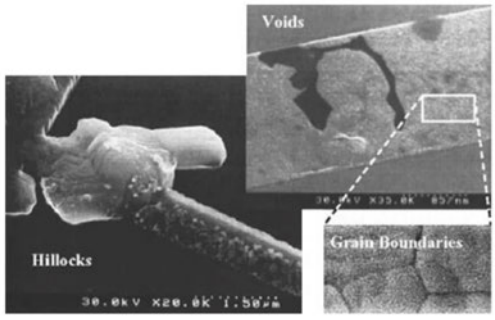

Fig. 5: Hillock and void formations in wires due to EM [29].

smaller feature size. Additionally, HCI-induced degradation can be recovered to a certain limit [28] and is affected by voltage, temperature and workload as shown in Figure 4.

3) Electromigration: Electromigration is a wear-out mechanism that occurs due to gradual movement of ions in a conductor (e.g. metal connections in a chip) due to momentum transfer between electrons and diffusing metal atoms. EM occurs when a high density current flows through wire over a long period of time. Such physical migration of metal ion (e.g. copper) can create a void or hillock that eventually results into an open or a short circuit causing a failure. Figure 5 shows the effect of EM on a copper interconnect [29].

All these aging and wear-out mechanisms greatly decrease device reliability, and eventually shorten chip-lifetime increasing possible failure rate as shown in Figure 6 [30]. Equations (1) to (3) show that degradation mechanisms also strongly depend on run-time conditions and workload. Although such mechanisms are quite slow, and the degradation magnitude is relatively hard to predict being statistical in nature, in most cases an accelerated aging (i.e. running the chip in a higher voltage and/or temperature than the nominal condition) allows us to determine the possible impact and lifetime, and deploy compensating mechanisms if possible.

D. Reliability vs. Security: The Good, The Bad, and The Ugly

We see that the transistor performance deviates from its nominal value due to process variation, environmental condition, and aging, as shown in Figure 7. Hence for designing high-performance circuits, we must develop solutions that minimize systematic and random process variations to an acceptable level, if not eliminated, by overcoming process technology challenges, and proposing robust and reliable design techniques that are tolerant to environmental variation and, to some extent, aging [3].

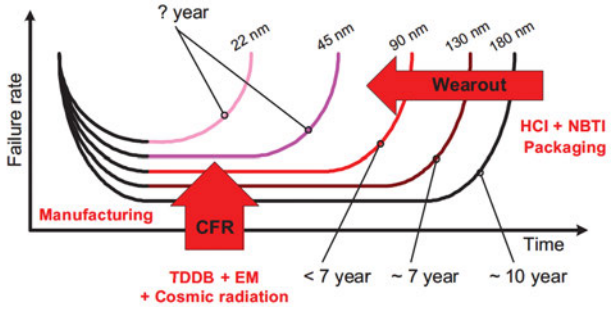

Fig. 6: Reliability bathtub variation with CMOS scaling [30].

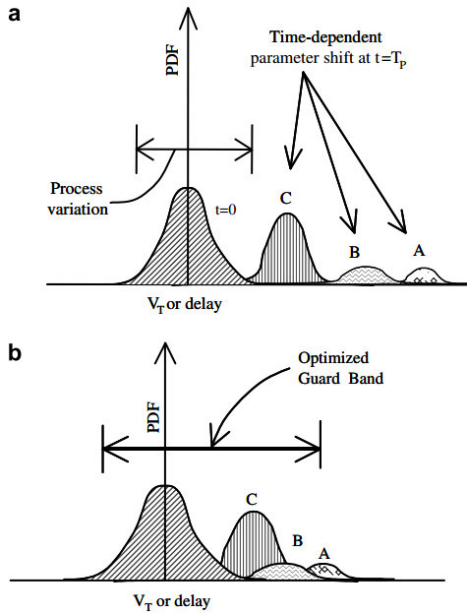

Fig. 7: (a) Transistors have different degrees of parameter drifts based on usages. (b) A redesigned model incorporating process variations and time-dependent shifts can lead to optimal performance [3].

However, these variations and degradation mechanisms do not necessarily have adverse impacts on hardware-oriented security primitives and applications. In fact, some of these variations and degradation mechanisms can be leveraged for ensuring hardware-based security. For example, PUFs and TRNGs rely on the manufacturing process variations, and an increase in such physical variations can potentially increase the PUF/TRNG outcome quality. In addition, detection of some types of counterfeit electronics can benefit from inherent aging and wear-out mechanisms, since presence of signs of prior usage can potentially lead to the detection of recycled chips.

However, these mechanisms are not always beneficial to all security applications either. Hence we would like to use the terms – *good*, *bad*, and *ugly* – to qualitatively state the relationship between process variations, reliability degradation and various hardware-based security mechanisms, as shown in Table I. The first column of Table I shows some conventional and security based applications and primitives, and respective rows refer how process variation, temperature, power supply noise, aging and wear-out mechanisms effect the quality of operation. Here, the *good* indicates that given variation or

TABLE I: Design and Technology Characteristics vs. Security Trade-off

Application /Primitive	Process Variation	Temper-ature	Power Supply Noise	Aging (BTI/ HCI)	Wear-out (EM)
Logic/ Memory Design	Ugly	Bad	Bad	Bad	Bad
PUF	Good	Bad	Bad	Bad	Bad
TRNG	Good	Bad	Bad	Bad	Bad
Recycled Electronics Detection	Ugly	Bad	Bad	Good	Good
Reliability Trojan Activation (In Field)	Bad	Bad	Bad	Ugly	Ugly
Reliability Trojan Detection (Prior to Operation)	Bad	Good/ Bad	Good/ Bad	Good	Good

degradation mechanism is actually desirable and beneficial to a security application or primitive; the *bad* means that it should be avoided if possible, and the *ugly* means that it is highly undesirable for a reliable operation. For example, in logic/memory applications, manufacturing "process variation" is highly undesirable (*ugly*) to ensure better performance, whereas it is one of the key requirements (*good*) for PUF and TRNG applications. "Aging" and "wear-out" mechanisms are *bad* for both regular logic/memory applications and PUFs, however they can be leveraged (*good*) for detecting recycled electronics.

In light of above discussion, it is crucial that we find a sweet spot between performance, reliability, and security. Hence, rather than building security primitives using logic/memory application oriented designs, and vice versa, the designers should make a balance of performance, reliability, and security to obtain the best trade-off depending on their target application. This will be further elaborated in subsequent sections.

IV. PUF AND TRNG

Both PUF and TRNG fundamentally share the concept that the process variation present in electronic devices can be leveraged to harness random entropy, although their requirements vary. In this section, we elaborate how process variation and reliability affect PUF and TRNG performance.

A. PUF

One can assess how good a PUF is for security applications using quality metrics – *uniqueness*, *randomness*, and *reliability*. *Uniqueness* measures distinctive quality of PUFs among one another based on their challenge-response pairs (CRPs), and *randomness* indicates the unpredictability of the PUF response. Additionally, *reliability* of a PUF assesses the capability to generate same CRPs over different environmental condition throughout its operational lifetime. It is crucial that a PUF maintains above qualities as close as possible to the ideal ones throughout its lifetime. Systematic process variation may create slightly biased output reducing uniqueness and randomness, whereas high random process variation increases the quality [31]. More importantly, environmental variations, such as power supply and temperature variations, and aging

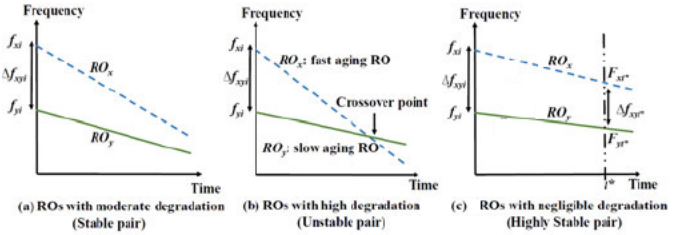

Fig. 8: Bitflip due to frequency degradation in RO-pair [32].

and wear-out mechanisms, generate unwanted error in PUF outputs making it unreliable for cryptographic applications. Some PUFs may produce upto 25% error in generated key due to aging [32].

To understand why PUF generates erroneous response, we revisit the conventional RO-PUF shown in Figure 1, and consider the frequency profile of a randomly selected RO-pair shown in Figure 8 [32]. For the given RO-pair, if the frequency of RO_x ring oscillator (f_{xi}) is greater than that of RO_y (f_{yi}), then '1' (otherwise '0') is generated as a response (Figure 8(a)). However it fails to generate a reliable (i.e. same as before) response if a crossover happens (i.e. $f_{xi} < f_{yi}$ after possible frequency degradation due to environmental variation and aging (Figure 8(b)). For maintaining maximum reliability, the two frequencies should never cross each other, maintaining a minimum frequency difference (i.e. frequency threshold (Δf_{th})) to compensate counter resolution if necessary, till the end of operational lifetime t^* (Figure 8(c)). Other PUF structures also suffer from similar reliability issues. Since it is crucial for a PUF to maintain reliable (i.e. error-free) response, researchers have proposed different techniques to minimize such errors:

1) Bootstrapping with Error Correction Code: A common solution for ensuring reliability is to use schemes like error correcting code (ECC) with PUF. ECC can generate reliable PUF output upto a certain margin despite presence of noise [33]. However, it relies on helper data that may partially reveal the secret key and potentially compromise the PUF. Most ECC schemes require redundant gates and an additional decode unit. Thus, ECC requires large area, power, and timing overheads making it impractical in resource constrained applications.

2) Aging Resistant Architecture: NBTI and HCI-aware aging resistant ARO-PUF was proposed in [32] as shown in Figure 9(a). This design has additional pull-up and pass transistors within the conventional RO-PUF architecture to reduce possible aging degradation. It has two modes of operation. At Oscillatory mode (EN = 1), it performs regular PUF operation (Figure 9(b)). It suffers from unavoidable AC Stress in this mode, however since the PUF activation time in the field is negligible, we can consider the degradation to be minimal. In the non-oscillatory mode when EN = 0 (Figure 9(c)), it removes DC stress for PMOS transistors as it ties them to V_{dd} to eliminate NBTI. It also breaks the RO chain, and removes AC (oscillatory) stress to eliminate HCI. So this design successfully minimizes aging degradation due to NBTI and HCI by eliminating stress when the PUF is not active.

The ARO-PUF shows promising result in terms of higher reliability (lower error) and lower frequency degradation. Figure 10 shows simulated frequency degradation profile for 5

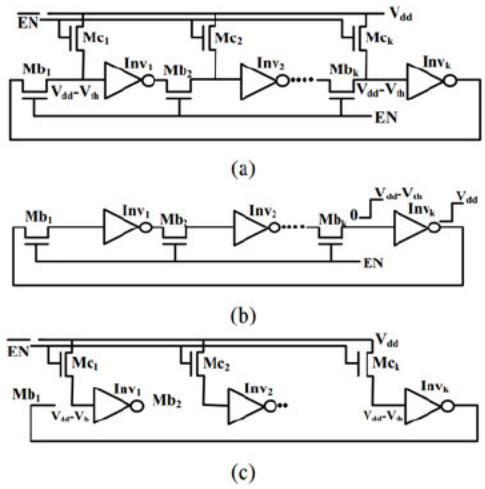

Fig. 9: (a) Aging-Resistant RO (ARO), (b) Oscillatory mode, and (c) Non-oscillatory mode [32].

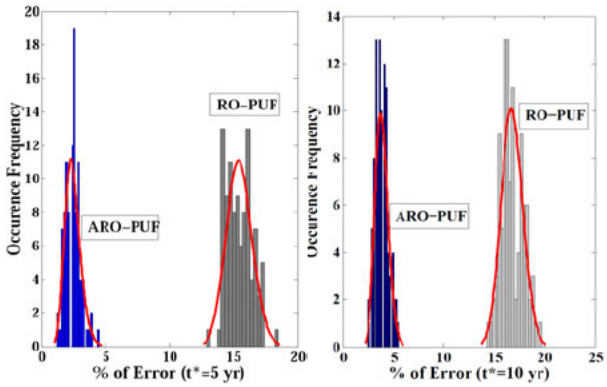

Fig. 10: Percentage of error after 5 and 10 years of operation for ARO-PUF and RO-PUF under same stress condition [32].

and 10 years of operational lifetime. Table II shows that ARO-PUF produces less error due to aging compared to RO-PUF.

3) Reliable RO-Pair Formation: Typically the bitflip in RO-PUF occurs due to high mismatch in frequency degradation rate in different ROs from environmental variations and aging. However, as discussed in Section III, degradation also depends on systematic and random variations and other parameters. Since randomly formed RO-pairs do not show any predictive correlation with such systematic and other variations within their elements, such a formation scheme tends to fail to achieve high reliability. A reliable pair formation scheme called RePa, proposed in [34], can achieve upto 100% reliability, i.e. zero error, by making selective RO-pairs. Using a predictive aging/voltage-based degradation profile, it sorts all ROs in a RO-PUF to find the most suitable pairs to probabilistically offer no bitflip, and does the PUF registration accordingly.

TABLE II: Average error (%) for ARO-PUF and RO-PUF.

Year	ARO-PUF	RO-PUF
2.5	1.5781%	10.3602%
5	2.4316%	11.4690%
7.5	3.1680%	12.1970%
10	3.8320%	12.7632%

TABLE III: Correlation of V_{dd} variation vs. aging degradation.

HSPICE Simulation			Silicon Result		
Stress (Year)	Correlation Coefficient (ρ)		Stress (Hour)	Correlation Coefficient (ρ)	
	Avg.	Worst		Avg.	Worst
2.5	0.54	0.4	7	0.41	0.37
5	0.57	0.44	14	0.46	0.41
7.5	0.61	0.45	21	0.51	0.44
10	0.67	0.49	28	0.54	0.47

TABLE IV: Effectiveness of RePa on RO-PUF against aging.

	HSPICE Simulation		Silicon Result	
Frequency Threshold (Δf_{th})	Required ROs (Avg.)	Error (Avg.)	Required ROs (Avg.)	Error (Avg.)
0 MHz	1x	1.25%	1x	0.96%
$0.5 * \Delta f_{th_{min}}$	1.52x	0.41%	1.7x	0.16%
$\Delta f_{th_{min}}$	1.82x	0%	1.94x	0%

The authors in [34] predicts the aging degradation using V_{dd} variation because both simulation and silicon data suggest that the degradation trend (rate/slope) for V_{dd} variation shows strong correlation to degradation rate for aging, as shown in Table III. This potentially eliminates the need of slow and costly burn-in test for aging prediction since the required data for pair formation now can be collected during the electrical testing phase. Table IV also shows RePa's effectiveness, which is upto 100%, for given frequency threshold (Δf_{th}).

B. TRNG

Typically, random numbers are generated by comparing two symmetric systems (devices) that possess some process variation and random inner noise to serve as an entropy source. A sampler/extraction unit is used to extract this entropy to digital data. However, variations in the inherent process, operating temperature, V_{dd}, and aging introduce systematic bias at TRNG output by infusing large asymmetry between the systems. It decreases TRNG quality since the generated output is no more random, rather somewhat predictable [35]. In addition, the TRNG is affected by limited process variation when inner random entropy source is not sufficient enough for suitable randomness and throughput. Older and mature technologies possess less process variation and provides less randomness. The randomness of TRNG in such technologies can become even worse under environmental variations and different aging mechanisms. This opens up variety of hardware-based attacks to TRNGs. For example, an attacker can vary V_{dd} and temperature away from the nominal condition, and intentionally bias the output to extract the 'predictable' bitstream [13]. A frequency injection attack on RO-based TRNGs can effect clock jitter working as the entropy source, and can guess the key, for example from a smart card, with minimal effort [16].

It is apparent that, a variation in run-time environmental condition and aging severely degrades the TRNG performance, and threatens the security that it offers. To combat this, researchers have proposed cryptographic hash functions, von Neumann corrector, and stream ciphers to be employed to the TRNG outputs to achieve more uniformity and statistical randomness. However, It reduces the throughput, and increases area and power overhead. Further, vendor agnostic TRNG design for FPGA is proposed in [36]. The authors in [13]

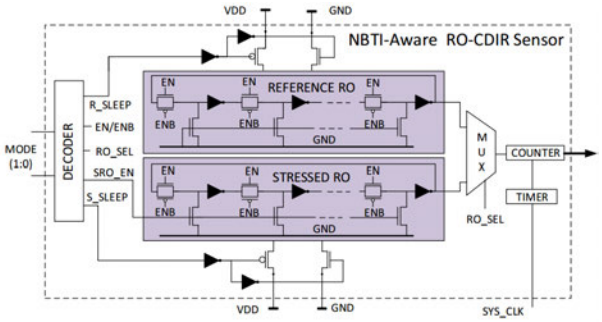

Fig. 11: NBTI-Aware RO-CDIR sensor [38].

proposed a technology-independent TI-TRNG to combat the security issues arising from such various degradation and variations. It uses a 'tunable'-RO architecture to leverage power supply noise along with clock jitter as the entropy source. This can overcome environmental variation and aging-induced bias by controlling jitter and adjusting RO delay by monitoring the run-time condition. A proposed power supply noise enhancement and tuning block, and a self-calibration scheme on bias detection further improves the performance, and serves against the hardware based attack.

V. Detecting Counterfeit Electronics

In today's complex electronic component supply chain, it is very challenging to detect and prevent the infiltration of counterfeit chips and FPGAs. As stated in Section II, a proper exploitation of electrical characteristics can lead to a cheaper, faster and more successful detection of counterfeit electronics. Since aging and wear-out mechanisms generally make a chip slower over time, one can estimate aging degradation of a circuit under test (CUT) by measuring its speed and comparing it with a reference speed from original unused (golden) chips. However, acquiring such reference (golden) measurements is not always feasible. In addition, manufacturing process variation and defects cause deviation in speed/delay or other electrical measurements even for golden chips. Hence it requires a large pool of golden data to maintain statistical significance. These issues are more prominent for legacy chips and COTS [19]. By exploiting such aging issues, researchers have proposed several techniques, such as embedding *design for anti-counterfeit* (DFAC) structures into the chip, or measuring degradation due to accelerated aging, for recycled IC detection.

A. Combating Die and IC Recycling

A combating die and IC recycling (CDIR) scheme takes aging into account to determine whether a chip has gone through a prior use. It is a lightweight and low-cost DFAC sensor with an RO-pair for self-referencing that eliminates the need for golden data [8], [37]. This RO-CDIR sensor contains two equal length ROs named reference-RO and stressed-RO, and compares their frequencies. Reference-RO is designed to age slowly, where the stressed-RO is designed to age at a much faster rate. When in operation mode, the stressed-RO's rapid aging reduces its speed (frequency), while the reference-RO's speed (frequency) largely remains the same. Thus, a large difference between the RO-frequencies implies that the chip has been used. A close physical placement of the ROs further reduces global and local process variations and environmental

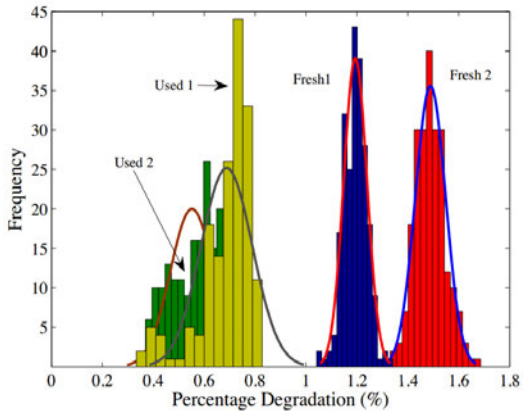

Fig. 12: Aging degradation distribution differences between used and fresh FPGAs. Here the used FPGA is aged for more time in the first case with different configuration. [39].

variations to give a finer measurement of the usage time. However, a limitation of this approach is that only half of the PMOS transistors experience DC NBTI stress, hence limited degradation, due to oscillatory nature of the scheme.

Figure 11 shows an NBTI-aware RO-CDIR sensor that exploits NBTI-induced degradation for an improved detection scheme [38]. While in operation mode, it gives maximum NBTI (DC) stress to the stressed-RO by breaking the RO chain and connecting all inverter inputs to ground so that they do not get a chance to recover from aging. However a partial recovery may occur when the chip is completely powered off. The stressed-RO's structure is mimicked by the reference-RO to avoid parametric variations. However, during operation, the reference-RO is kept disconnected from the power and ground line to minimize aging. Since the two ROs have different aging stress, their frequencies continues to deviate over time, and it increases the probability of a more accurate detection.

B. Recycled FPGA Detection

FPGA is widely use in various regular and critical applications. Since it provides flexibility and reconfigurability, its characteristics highly vary from vendor to vendor and on applications. This makes the detection of recycled FPGAs a very challenging task.

The authors in [39] showed a two step electrical test based recycled FPGA detection scheme. The key idea is to exploit the performance variation between new and used FPGAs when the same stress conditions (high voltage, high temperature, and transistor switching) are applied. Since FPGA can be reconfigured unlike ASIC, one can implement on-chip test structures, such as ROs, to measure the aging degradation. Figure 12 shows some interesting findings of this work:

- Fresh and previously used (recycled) FPGAs experience different degradation rate even if they face similar aging stress. Eventually the degradation saturates with aging, hence fresh FPGAs experience rapid (and more) degradation compared to used FPGAs for a particular aging cycle or time.
- New FPGAs have a tighter degradation profile compared to that of used FPGAs. It is because the newly configured designs in fresh FPGAs only experience process variations,

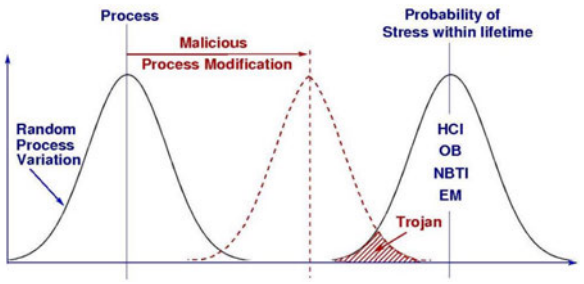

Fig. 13: Shifting the Process Variation to induce a Reliability Trojan based on Aging Stress [40].

and slight environmental variation, whereas the used FPGAs have already suffered from multiple variation sources.

- Although the degradation profile varies based on aging condition and implemented circuit, the basic degradation trends for new and used FPGAs remain the same. For example, in Figure 12, both used FPGAs show lower degradation with a larger variance compared to fresh ones, even though former two actually experienced different initial aging stress.

Based on these characteristics, the authors in [39] used a support vector machine (SVM)-based classifier to perform a two step filtering mechanism for detecting any prior use. Phase 1 detects the used FPGAs that are directly classifiable via SVM using the golden data. Phase 2 performs a quick precision acceleration on all FPGAs and finds the ones that degrades more. This data is fed to classifier to finally detect used (recycled) FPGAs.

We see that aging degradation is highly beneficial for detecting recycled ICs and FPGAs. However, in most techniques, a reference is needed against which the degradation is typically compared. However, based on the findings from [39], and modeling necessary classifiers with possible self referencing techniques, one can leverage aging degradation for counterfeit (recycled) electronics detection.

VI. Reliability-based Hardware Trojans

We see from Section II that hardware Trojans can have numerous forms and possibilities in terms of attack (insertion), and detection and prevention of these Trojans are extremely difficult. To remain in the spirit of this paper, we focus on *process reliability-based hardware Trojans* only, since it leverages process variation and reliability issues for attack and even for possible detection mechanisms [40], [41].

A process reliability based Trojan reduces the reliability of the chip through malicious alterations of the manufacturing process conditions. It does not necessarily require additional circuitry, or alter functionality of the original design. Such a reliability Trojan leverages aging degradation and wear-out mechanisms, or utilizes other manufacturing process and runtime variations. A minor shift of the critical parameters in manufacturing process, such as increase of random dopant fluctuations, forced defect generation, insufficient passivation, and accelerated aging, can introduce such Trojans that reduces the reliability of the system to a point of premature failure, as shown in Figure 13 [40]. A reliability Trojan is extremely difficult to detect, and the resultant premature failure might lead to catastrophic effects in the field.

A. Reliability Trojan Insertion

There are various options for placing reliability Trojans inside a chip. An attacker can alter individual transistors in the critical path, can design a reliability Trojan in gate level, or even place it in layout. Such a Trojan insertion can be coupled with design level changes, such as voltage or temperature variations, for a cumulative effect since these degradation parameters have strong dependency on environmental conditions as well.

1) Transistor-level Reliability Trojan: An attacker can choose some individual transistors, or a portion of the critical path, where he/she makes the malicious modification in the critical manufacturing process parameters in a way such that it causes a premature failure, or creates a backdoor to leak critical data. It does not require any additional gates, or processing, and the Trojan may only be detected when the circuit is compromised. The authors in [41] proposed a stealthy dopant-level hardware Trojan which can be implemented by changing the dopant polarity of existing transistors. Since the modified circuit appears legitimate on all wiring layers (including all metal and polysilicon), and only silicon active layer is affected, it becomes very difficult to detect. In addition to possible functional failure, a clever use of this Trojan can leak critical data via side-channel.

2) Gate-level Reliability Trojan: The attacker can carry out this attack in addition with traditional gate level Trojans, or can perform this separately using dummy/additional gates/designs. For example, we consider a ring-oscillator whose frequency can be measured and compared against a reference value to generate a trigger that consequently compromises the system security. The attacker alters the RO transistors to act as an aging-based reliability Trojan that experiences rapid degradation due to additional defects and/or stress. The frequency degradation of the RO Trojan becomes faster, and eventually it becomes less than the reference value. This can trigger the system to carry out an unauthorized process. For example, such an attack poses high vulnerability for the systems that use on-chip ROs for self-referencing and compensation.

3) Layout-Level Reliability Trojan: The attacker can use wear-out mechanisms such as electromigration to carry out such attack. In this case the attacker intentionally thins out the metal contact than it should be, and may introduce additional defects. Due to higher current density, EM would be faster in this altered connection than that of the nominal design. Hence the connection would experience a faster failure creating a permanent open connection.

B. Reliability Trojan Detection

Detection of various reliability Trojans is extremely difficult since their effects may be overshadowed by the process and environmental variations. If an IC with a reliability Trojan is already in operation, then the Trojan would eventually activate, and would act as a probabilistic 'time-bomb' based on the failure mechanism and degradation rate. In such a case, process and environmental variations act mostly as a promoter to accelerate the failure mechanism as given by Equations (1) to (3). However, if the reliability Trojan is still not activated, i.e. the detection scheme is implemented prior to operation, then the environmental degradation can be used in favor of detection via accelerated aging. In such a case, a

number of samples undergo accelerated aging test, and their performance degradation is evaluated. If the chip contains a reliability Trojan, accelerated aging may then activate it. Based on the discussion given in Section III-C, an accelerated aging would show a larger degradation for the Trojan-inserted chip compared to the authentic one, and hence it would possess a higher probability of detection.

VII. Emerging Nanodevices: Reliability Challenges and Security Opportunities

Most of the security primitives and applications that exist today heavily rely on CMOS technology. However, with the maturity of CMOS technology, leveraging some of the security aspects (such as extracting entropy from random variation as in a TRNG) is getting tighter for older nodes. Further, current CMOS-based security primitives/countermeasures may seem inadequate with new emerging threats and vulnerabilities. Nanoscale devices and technologies such as phase-change memory (PCM), memristors, and carbon nano-tubes (CNTs) can offer promising applications in such regards. Since these devices are not yet mature enough, in many cases they exhibit unorthodox properties that may be leveraged for security. We take PCM and CNTs as examples to show the challenges and opportunities that these emerging devices may offer.

A. Phase Change Memory

PCM is an emerging nanoscale device that enables non-volatile storage with high density and fast read/write operations. The reset/set operation for memory application is based on the transition to-and-from an amorphous (high resistance) phase and a crystalline (low resistance) phase with a resistance difference of around 10^4-ohm between the two phases [42]. Some interesting properties of PCM from the security point of view are:

- *Programming Variability and Random Telegraph Noise (RTN):* PCM cells show stochastic programming variability. For example, given two PCM cells, a reset operation on them with the same reset pulse yields two close but different resistance values [42]. PCM also exhibits RTN that occurs in PCM devices as short-term resistance fluctuations, with power spectral density varying with PCM cell contact area, temperature, applied voltage, etc.
- *Resistance Drift:* Due to resistance drift an amorphized PCM cell may have an increase or 'drift' in resistance over time, and eventually change to crystalline phase with a drastic decrease in resistance [43].

While these properties may be considered as problematic issues for reliable read/write and data retention, it potentially has interesting and useful applications in security:

- *Anti-Counterfeiting Sensors:* Resistance drift in PCM cells could be used to design passive aging-sensors for time keeping or time stamping, and to detect how long an IC or electronic system has been in the supply chain. Such a passive sensor does not need to be powered on for detecting the age of the chip/system, and hence has advantages over power-based timers.
- *TRNG:* PCM has several inherent entropy sources such as RTN and structural variability. In addition, the programming variability that PCM exhibits can potentially be leveraged to build TRNG. However, the random variability and other attack scenarios are yet to be assessed.

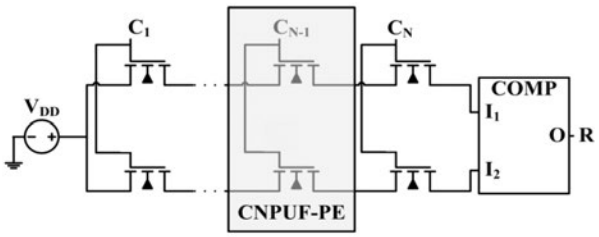

Fig. 14: CNPUF proposed in [45]. Characteristics of CNPUF parallel element varies due to process variation.

B. Carbon NanoTube

CNT and CNT based field effect transistors (CNTFETs) have interesting properties governed by their unique structure. It also offers a platform to potentially integrate digital logic with non- logic components such as analog circuitry and sensors [44]. Although, CNT and CNTFETs still lack maturity and suffers high yield problem, some of its 'not-so-mature' yet unique properties that may have interesting applications in developing hardware security primitive are as follows:

- *Variability:* The property of CNTFETs largely depend on the channel CNT properties (e.g., semiconducting or metallic, etc.), length and patterning, contact, etc. Hence the inherent sources of variability are quite large and heavily depend on manufacturing processes.
- *Channel Sensitivity:* The channel material in CNTFETs is highly sensitive to carrier mobility variation with operating conditions, channel contamination, physical deformation in channel nanotubes, by photons, and other phenomena. Hence, the issue of controlling channel quality has received much attention for maintaining reliability.

It is apparent that the variability and low reliability that CNT shows are not properly suitable for logic applications. However, higher process variation makes CNTs and CNTFETs intriguing candidates for building PUFs. In [45], a carbon nanotube based PUF (namely CNPUF, see Figure 14) was proposed. It leverages the fact that the lack of chirality control in the manufacturing process yields metallic CNTs over semiconducting CNTs in a non-deterministic way. Utilizing the characteristic variation between semiconducting and metallic properties of CNTs can lead to distinguishable but random states since the off-current for semiconducting CNTs is considerably lower than that of metallic CNTs. Such inherent random properties can be leveraged to build nanoscale PUFs.

We see that the emerging nano-devices such as PCM and CNT have promising security potentials, even though they suffer from high variability. That brings us back to the fundamental question: what should be the trade-off for performance, reliability and security? Finding a sweat spot in between them eventually has become essential, so that the designers can enjoy a high performance with improved security.

VIII. Conclusion

In this paper we have presented challenges and opportunities between performance, reliability and security. In contrast to traditional performance-oriented designs, we have explored

the security opportunities leveraging the sources associated with manufacturing process variations, reliability degradation and aging. We see that there are many security applications that utilize random manufacturing variations and aging. We also point out some of the cases where both performance and security suffer from same issues. We conclude that there is no one solution that works efficiently for all objectives; rather we need to consider them in a holistic way to find a sweet spot between performance, reliability and security. We also see that some emerging nanoscale devices can potentially provide us some security benefits, even though they might lack maturity in terms of performance.

REFERENCES

[1] K. J. Kuhn et al "Process technology variation," *IEEE Transactions on Electron Devices*, vol. 58, no. 8, pp. 2197–2208, 2011.

[2] C. Kenyon et al., "Managing Process Variation in Intel's 45nm CMOS Technology," *Intel Technology Journal*, vol. 12, pp. 93–110, 2008.

[3] M. Alam, "Reliability and process-variation aware design of integrated circuits," *Microelectronics Reliability*, vol. 48, pp. 1114–1122, 2008.

[4] M. Tehranipoor and C. Wang, *Introduction to hardware security and trust*, Springer, 2011.

[5] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld, "Physical one-way functions." *Science*, vol. 297, no. 5589, pp. 2026–30, sep 2002.

[6] B. Gassend, D. Clarke, M. Van Dijk, and S. Devadas, "Silicon physical random functions," in *Proceedings of the 9th ACM conference on Computer and communications security*, p. 148, nov 2002

[7] B. Jun and P. Kocher, "The intel random number generator," *Cryptography Research Inc. white paper*, 1999.

[8] X. Zhang and M. Tehranipoor, "Design of On-chip Lightweight Sensors for Effective Detection of Recycled ICs," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 22, pp. 1016–1029, 2014.

[9] M. Tehranipoor and F. Koushanfar, "A Survey of Hardware Trojan Taxonomy and Detection," *IEEE Design and Test of Computers*, vol. 27, no. 1, pp. 10–25, 2010.

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

[11] G. E. Suh and S. Devadas, "Physical Unclonable Functions for Device Authentication and Secret Key Generation," in *Proceedings of the 44th annual Design Automation Conference*, ACM, pp. 9–14, 2007.

[12] J. Guajardo, B. Škorić, P. Tuyls, S. S. Kumar, T. Bel, A. H. M. Blom, and G. J. Schrijen, "Anti-counterfeiting, key distribution, and key storage in an ambient world via physical unclonable functions," *Information Systems Frontiers*, vol. 11, no. 1, pp. 19–41, 2009.

[13] M. T. Rahman, Kan Xiao, D. Forte, Xuhei Zhang, J. Shi and M. Tehranipoor,"TI-TRNG: Technology independent true random number generator", *Design Automation Conference (DAC), 2014 51st ACM/EDAC/IEEE*, San Francisco, CA, 2014, pp. 1-6.

[14] A. Rukhin et al., "A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications ," DTIC Document, Tech. Rep., 2001.

[15] W. Schindler and W. Killmann, "Evaluation Criteria for True (Physical) Random Number Generators Used in Cryptographic Applications," in *Cryptographic Hardware and Embedded Systems-CHES 2002*, Springer, 2002, pp. 431–449.

[16] A. T. Markettos and S. W. Moore, "The Frequency Injection Attack on Ring-Oscillator-Based True Random Number Generators," in *Cryptographic Hardware and Embedded Systems-CHES 2009*, Springer, 2009, pp. 317–331.

[17] U. Guin, D. DiMase, and M. Tehranipoor, "Counterfeit integrated circuits: detection, avoidance, and the challenges ahead," *Journal of Electronic Testing*, vol. 30, no. 1, pp. 9–23, 2014.

[18] S. Shahbazmohamadi, D. Forte, and M. Tehranipoor, "Advanced physical inspection methods for counterfeit ic detection," in *ISTFA 2014: Conference Proceedings from the 40th International Symposium for Testing and Failure Analysis*, ASM International, 2014, p. 55.

[19] M. M. Tehranipoor, U. Guin, and D. Forte, *Counterfeit Integrated Circuits: Detection and Avoidance*, Springer, 2015.

[20] R. Kumar, "Interconnect and noise immunity design for the Pentium 4 processor," *Proc. of Design Automation Conference*, pp. 1–12, 2003.

[21] Y. Wang et al., "A 4.0 GHz 291 Mb voltage-scalable SRAM design in a 32 nm high-k + metal-gate CMOS technology with integrated power management," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 1, pp. 103–110, 2010.

[22] R. Kumar and V. Kursun, "Reversed temperature-dependent propagation delay characteristics in nanometer CMOS circuits," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 53, no. 10, pp. 1078–1082, 2006.

[23] E. Maricau and G. Gielen, *Analog IC Reliability in Nanometer CMOS*, Springer, 2013.

[24] S. Zafar et al."A comparative study of nbti and pbti (charge trapping) in sio2/hfo2 stacks with fusi, tin, re gates," in *VLSI Technology, 2006. Digest of Technical Papers. 2006 Symposium on*, IEEE, 2006, pp. 23–25.

[25] A. Tiwari and J. Torrellas, "Facelift: Hiding and slowing down aging in multicores," *Proceedings of the Annual International Symposium on Microarchitecture, MICRO*, pp. 129–140, 2008.

[26] W. Wang et al., "The impact of NBTI on the performance of combinational and sequential circuits," *Proceedings - Design Automation Conference*, pp. 364–369, 2007.

[27] D. Lorenz, M. Barke, and U. Schlichtmann, "Aging analysis at gate and macro cell level," *IEEE/ACM International Conference on Computer-Aided Design, Digest of Technical Papers, ICCAD*, pp. 77–84, 2010.

[28] D. Saha, D. Varghese and S. Mahapatra, "On the generation and recovery of hot carrier induced interface traps: a critical examination of the 2-D R-D model," in IEEE Electron Device Letters, vol. 27, no. 3, pp. 188-190, March 2006.

[29] J. Lienig and G. Jerke, "Current-driven wire planning for electromigration avoidance in analog circuits," in *Proceedings of the 2003 Asia and South Pacific Design Automation Conference*, ACM, 2003, pp. 783–788.

[30] B. Mesgarzadeh, I. S. Saab, and A. Alvandpour, "Reliability challenges in avionics due to silicon aging," *Proceedings of the 2012 IEEE 15th International Symposium on Design and Diagnostics of Electronic Circuits and Systems, DDECS 2012*, pp. 342–347, 2012.

[31] A. Maiti, V. Gunreddy, and P. Schaumont, "A systematic method to evaluate and compare the performance of physical unclonable functions," *Embedded Systems Design with FPGAs*, pp. 245–267, 2013.

[32] M. T. Rahman, F. Rahman, D. Forte, and M. Tehranipoor, "An Aging-Resistant RO-PUF for Reliable Key Generation," *IEEE Transactions on Emerging Topics in Computing*, vol. PP, no. 99, pp. 1–1, 2015.

[33] M. D. Yu and S. Devadas, "Secure and Robust Error Correction for Physical Unclonable Functions," *IEEE Design & Test of Computers*, vol. 27, no. 1, pp. 48-65, Jan.-Feb. 2010.

[34] M. T. Rahman, D. Forte, F. Rahman, and M. Tehranipoor, "A Pair Selection Algorithm for Robust RO-PUF against Environmental Variations and Aging," *IEEE International Conference on Computer Design*, pp. 415–418, 2015.

[35] B. Sunar, W. Martin, and D. Stinson, "A Provably Secure True Random Number Generator with Built-In Tolerance to Active Attacks," *IEEE Transactions on Computers*, vol. 56, no. 1, pp. 109–119, 2007.

[36] D. Schellekens, B. Preneel, and I. Verbauwhede, "FPGA vendor agnostic true random number generator," *Proceedings - 2006 International Conference on Field Programmable Logic and Applications, FPL*, pp. 139–144, 2006.

[37] X. Zhang, N. Tuzzio, and M. Tehranipoor, "Identification of recovered ics using fingerprints from a light-weight on-chip sensor," in *Proc. of Design Automation Conference*, ACM, 2012, pp. 703–708.

[38] U. Guin, X. Zhang, D. Forte, and M. Tehranipoor, "Low-cost On-Chip Structures for Combating Die and IC Recycling," *Proceedings of the The 51st Annual Design Automation Conference on Design Automation Conference - DAC '14*, pp. 1–6, 2014.

[39] H. Dogan, D. Forte, and M. M. Tehranipoor, "Aging analysis for recycled FPGA detection," *Proceedings - IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems*, no. February 2016, pp. 171–176, 2014.

[40] Y. Shiyanovskii et al., "Process reliability based trojans through NBTI and HCI effects," *2010 NASA/ESA Conference on Adaptive Hardware and Systems, AHS 2010*, pp. 215–222, 2010.

[41] G. T. Becker, F. Regazzoni, C. Paar, and W. P. Burleson, "Stealthy dopant-level hardware trojans," in *Cryptographic Hardware and Embedded Systems-CHES 2013*, Springer, 2013, pp. 197–214.

[42] H. P. Wong et al, "Phase change memory," *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2201–2227, 2010.

[43] F. Dirisaglik et al., "High speed, high temperature electrical characterization of phase change materials: metastable phases, crystallization dynamics, and resistance drift," *Nanoscale*, vol. 7, no. 40, pp. 16 625–16 630, 2015.

[44] A. D. Franklin,"Carbon nanotube electronics," Emerging Nanoelectronic Devices, ed. A. Chen, John Wiley & Sons, Ltd, Jan. 2015.

[45] S. Konigsmark, L. K. Hwang, D. Chen, and M. D. Wong, "Cnpuf: A carbon nanotube-based physically unclonable function for secure low-energy hardware design," in *Design Automation Conference (ASP-DAC), 2014 19th Asia and South Pacific*, IEEE, 2014, pp. 73–78.