Pseudo-Chaotic Lossy Compressors for True Random Number Generation

Tommaso Addabbo, Ada Fort, Santina Rocchi, Valerio Vignoli

Information Engineering Department
University of Siena, 53100 Italy
(e-mail: addabbo@dii.unisi.it)

Ljupco Kocarev

Academy of Science and Arts of Skopje, Macedonia

⋆ RESEARCH MANUSCRIPT ⋆

– PLEASE REFER TO THE PUBLISHED PAPER[1] –

Abstract

This paper presents a compression method that exploits pseudo-chaotic systems, to be applied to True Random Bit Generators (TRBGs). The theoretical explanation of the proposed compression scheme required the projection of some results achieved within the Ergodic Theory for chaotic systems on the world of digital pseudo-chaos. To this aim, a weaker and more general interpretation of the Shadowing Theory has been proposed, focusing on probability measures, rather than on single chaotic trajectories. The design of the compression scheme has been theoretically discussed in order to assure the final entropy of the compressed TRBG to be arbitrarily close to the maximum limit of 1 bit/time-step. The proposed solution requires extremely low-complex hardware circuits for being implemented, assures a constant throughput and is based on theoretical results of general validity.

1 Introduction

Since the pioneering work of Shannon, methods for processing information are based on certain modeling of information sources (e.g., ergodic processes) [1]. Once the stochastic model is known, a processor (e.g., a compression algorithm) can be optimally designed. Otherwise, the processing is achieved on the basis of the observed data only, adopting sub-optimal solutions that must comply with some complexity constraints. In both cases, the information processors are typically aimed at preserving some level of information intelligibility: for

[1]Circuits and Systems I: Regular Papers, IEEE Transactions on, 2011, vol. 58, n. 8, p. 1897-1909.

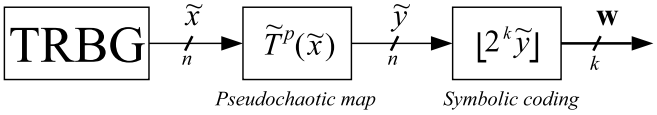


Figure 1: The block diagram illustrating the compression scheme ($k < n$).

example, when a digital photo is coded and compressed, an adequate perceptive recognizability of the subject is somehow supposed to be guaranteed [1].

A different point of view must be adopted when dealing with cryptographic True Random Bit Generators (TRBGs). In such cases, the target is the *theoretical unpredictability*[2] of the source, i.e., one aims at precluding any theoretical possibility for an observer to exploit forms of statistical biasing of the past data to predict the future bits.

Even if several methods have been proposed in literature to achieve adequate levels of practical unpredictability of a source [2–7], the only theoretically-proven way to better approximate an unpredictable fair coin tosser with a TRBG is to apply optimal compression to its output, boosting its entropy. The problem of compressing data has been widely investigated, and several methods have been proposed in literature for which some optimality criteria were assessed theoretically [1, 8]. The main drawback of these effective methods is the involved complexity: regardless of the adopted approach (e.g., entropy coding or adaptive variable coding) quite complex algorithms and circuits have to be implemented to obtain the compression result [9–11]. By renouncing the preservation of data intelligibility, less complex compression methods were proposed in literature. Von Neumann in 1951 proposed a maximum entropy lossy compressor with not-constant output bit-rate, working for sequences of i.i.d. random variables [12]. Several years later other authors generalized the same approach for finite-memory Markovian sources, involving algorithmic complexities that grow exponentially with the depth of the process memory [13, 14]. As far as simpler solutions are addressed, under-sampling methods have to be taken into account. These methods are valid for generic binary ergodic sources with vanishing statistical dependencies. In such case the TRBG sequence is decimated, obtaining a binary source with constant bit-rate, with an entropy that can grow up to an asymptotic value, that is equal to the entropy of a Bernoulli-distributed binary source, which in practical cases is always unbalanced.

In this paper, adopting a deep-lossy compression approach, we propose a solution that exploits pseudo-chaotic systems. The working of the proposed method is very simple, and referring to Fig. 1 it can be summarized in the following three steps:

1) collect n bits from the TRBG, to define the initial condition $\tilde{x}$ of a n-bit finite state machine implementing a pseudo-chaotic map $\tilde{T}$;
2) digitally compute $\tilde{T}^p(\tilde{x})$;
3) output the k most significant bits of the result, with $k < n$.

The ratio $\frac{k}{n}$ represents the compression ratio, and in this paper we prove the-

[2] We distinguish between theoretical unpredictability and practical unpredictability. The latter holds when predicting techniques are theoretically possible, but practically unfeasible [2–4].

oretically how to design the overall system to assure the final binary source to have an entropy greater than a given worst-case value.

To explain why and how this simple scheme works it is necessary to project some results achieved within the Ergodic Theory (valid for chaotic systems) on the world of digital pseudo-chaos. To this aim, we have proposed a weaker and more general interpretation of the Shadowing Theory proposed by Coomes et al. [15], focusing on probability measures, rather than on single chaotic trajectories. As it is made clearer throughout the paper, even if we discuss a specific example based on the Rényi chaotic map, the results proved in this work are general, and adopting the proposed design approach the general compression scheme of Fig. 1 can be adapted to other pseudo-chaotic maps, provided to prove the fulfillment of some theoretical assumptions. We have chosen the Rényi chaotic map since it is a not continuos map and its pseudo-chaotic implementation involves very efficient hardware circuits, as it has been recently technically demonstrated even for high digitization accuracy (i.e., with n ranging up to 10^3) [16]. This property allows us to work with our weaker and more general interpretation of the Shadowing Theory proposed by Coomes et al., that holds for continuos maps.

This paper is organized as in the following. In the next three Sections we discuss the preliminaries that are used to prove the main theoretical result, discussed in Sec. 5, and the design of the pseudo-chaotic compressor. In detail, in Sec. 2 we introduce a general result about the entropy of binary sources, that represents the design key for the proposed compression scheme. In Sec. 3 we introduce a conceptual link between binary sources and discrete-time real-valued stochastic processes. We exploit this link throughout the paper, to achieve the basic idea of the post-processing machine: the use of the mixing capabilities of chaotic maps to aim at a source of i.i.d. binary words, with almost uniform distribution. According to this approach, in Sec. 4 we analyze theoretically the use of a chaotic map to process a sequence of random variables. Since the final goal of this paper is to propose a digital post-processing machine, in Sec. 5 we propose a weaker and more general interpretation of the Shadowing Theory proposed by Coomes et al. [15], focusing on probability measures, rather than on single chaotic trajectories. This approach allows to analyze the effects of discretizing the chaotic domain when computing the map with finite arithmetic precision. In Secs. 6 and 7 we discuss a compression scheme based on the Rényi chaotic map, also applying the proposed solution to an actual TRBG circuit prototype. For a better readability of the paper, some proofs were reported in the Appendices.

2 Preliminaries: The entropy and its estimation

In this paper we assume TRBGs as ergodic sources of random bits, issuing one single bit at each time-step. In actual TRBGs bias and correlation between symbols deteriorate the statistical characterization of the generated sequences, and Information Theory provides theoretical tools for measuring how well the source approximates the behavior of an unpredictable fair coin thrower.

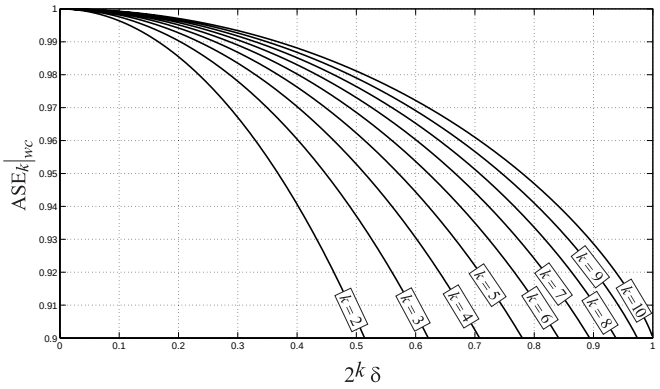


Figure 2: The worst case partial entropy as a function of the maximum allowed normalized deviation $2^k\delta$ for the terms $P(\beta)$ in eq. (2), for different values of k.

Definition 1 *The Average Shannon Entropy (ASE) of a TRBG is equal to*

$$ASE = \lim_{k\to\infty} -\frac{1}{k} \sum_{\beta\in\{0,1\}^k} P(\beta)\log_2 P(\beta), \tag{1}$$

where the summation extends to the finite set collecting the 2^k binary k-tuples of the form $\beta = \{b_0, \ldots, b_{k-1}\}$, $b_i \in \{0, 1\}$. The ASE indicates, for a given TRBG, the average amount of information issued at each time-step and therefore the result of (1) is expressed in bit/time-step.

It is immediate to check that for the ideal TRBG the ASE is equal to 1 bit/time-step, i.e., it is equal to the maximum Shannon Entropy for a binary source. The computation of (1) is unfeasible for most TRBGs, since the statistical distributions of any order are involved. Accordingly, the practical approach consists in truncating the expression (1) to a given integer $k > 0$, obtaining the Partial Average Shannon Entropy

$$\mathrm{ASE}_k = -\frac{1}{k} \sum_{\beta\in\{0,1\}^k} P(\beta)\log_2 P(\beta). \tag{2}$$

The above truncated expression will be used in the following, in order to simplify the analysis.

We conclude this section reporting an useful theorem valid for any ergodic binary source, proved in [17], that links the uniformity-degree of the probability mass function for the words β and the worst-case partial entropy of the source. In detail, let us assume for a TRBG that the generation probabilities for the k-bit words deviate from their ideal value $\frac{1}{2^k}$ by a quantity not greater than $\delta < \frac{1}{2^k}$, i.e.,

$$\left| P(\beta) - \frac{1}{2^k} \right| < \delta < \frac{1}{2^k}. \tag{3}$$

Accordingly, by defining $P_L = \frac{1}{2^k} - \delta$ and $P_H = \frac{1}{2^k} + \delta$, we have the following

Theorem 1 (Addabbo et al. [17]) *The partial entropy (2) satisfies the inequality*

$$ASE_k \geq -\frac{2^k}{2k}\left(P_L \log_2 P_L + P_H \log_2 P_H\right) = ASE_k|_{wc}. \tag{4}$$

In other words, the above theorem assures that a sufficient condition for a binary source to have a partial entropy greater than the worst-case $\mathrm{ASE}_k|_{wc}$ is to have a probability mass function in which each term $P(\beta)$ satisfies condition (3). Accordingly, the inequality (4) can be used for setting the maximum deviation δ that each term $P(\beta)$ is allowed to have from the uniform level $\frac{1}{2^k}$, such to have a partial entropy not lower than a worst-case lower bound (see Fig. 2). In order to express δ as a function of the $\mathrm{ASE}_k|_{wc}$, the eq. (4) can be inverted either numerically or referring to the following approximation, in which a 4th-degree truncated Taylor polynomial expression of $x \log_2 x$ was adopted:

$$\delta \approx \sqrt{\frac{3}{2^{2k}}\left(\sqrt{1+\frac{4k\ln 2}{3}(1-\mathrm{ASE}_k|_{wc})}-1\right)}. \tag{5}$$

The approximation implies a relative error lower than 4% on the actual $\mathrm{ASE}_k|_{wc}$ related to δ, for k ranging between 2 and 64, for $\mathrm{ASE}_k|_{wc} \geq 0.95$ bit/time-step. For example, the above approximation for $k = 8$ and $\mathrm{ASE}_k|_{wc} = 0.95$ yields $\delta \approx 2.79 \cdot 10^{-3}$, to which corresponds using (4) an actual $\mathrm{ASE}_k|_{wc} \approx 0.949$, involving a relative error $\approx 0.12\%$.

3 Preliminaries: Binary Sources as Quantized Representatives of Real-Valued Stochastic Processes

In this section we introduce those concepts that represent the foundations of this paper, by providing a theoretical interpretation of the n-bit variable obtained at the step 1) of the compression algorithm presented in the Introduction.

When n bits at a time are collected from a generic TRBG, a sequence of n-bit words $\{\mathbf{w}_i \in \{0,1\}^n, i \in \mathbb{N}\}$ is obtained. According to this point of view, the processed source is an information source $\mathbf{B}_n$ issuing n-bit words for which we assume the stationary joint probability distributions of any order well defined, and in general unknown (the stationarity of $\mathbf{B}_n$ derives from the ergodicity of the TRBG). In this paper we interpret theoretically the sequence of n-bit words as the result of a n-bit *quantization* of a *hypothetical* stochastic process $\mathbf{X} = \{x_i \in [0,1); i \in \mathbb{N}\}$. We stress that this hypothetical process has to be distinguished from that one actually used for generating the random bits, since it is a theoretical abstraction. As it is made clearer in the following, the existence of such an hypothetical process is assured by construction, since it can be theoretically constructed on the basis of the joint probability distributions associated to the n-bit word process.

Within the above theoretical framework, the generic word $\mathbf{w} = b_{n-1} \ldots b_0$ can be viewed as the binary label assigned to one of the 2^n disjoint intervals that evenly divide the domain $[0,1)$. From this point of view, we denote with

Table 1: Transition probabilities for the Markovian binary source $\mathbf{B}_2$.

	Next word generation probability			
Previous word	**00**	**01**	**10**	**11**
00	1/3	1/3	1/3	0
01	1/2	1/2	0	0
10	0	1/3	1/3	1/3
11	0	1/2	1/2	0

$\mathbf{j} = b_{n-1} \dots b_0$ the binary representation of the integer $0 \leq j < 2^n$, and with $\mathcal{P}_n = \{I_0, \dots, I_{2^n-1}\}$ the quantization partition made of 2^n equal intervals that divide $[0, 1)$. Therefore, a sequence $\{\mathbf{w}_i \in \{0, 1\}^n, i \in \mathbb{N}\}$ can be viewed as the n-bit quantization of any sequence $\{x_i \in [0, 1); i \in \mathbb{N}\}$ such that

$$\mathbf{w}_i = \mathbf{j} \Leftrightarrow x_i \in I_j.$$

According to the above expression, the quantization is performed by truncation. Form a probabilistic point of view, the only constraint that relates the binary process $\mathbf{B}_n$ and the hypothetical process $\mathbf{X}$ is that for any order joint probability it must result

$$P(\mathbf{w}_0 = \mathbf{j}_0, \mathbf{w}_1 = \mathbf{j}_1, \dots) = P(x_0 \in I_{j_0}, x_1 \in I_{j_1}, \dots). \tag{6}$$

For a given binary source $\mathbf{B}_n$, the set $\Omega(\mathbf{B}_n)$ of hypothetical stochastic processes satisfying (6) is infinite, and $\mathbf{B}_n$ can be defined as the *quantized representative* of those processes belonging to $\Omega(\mathbf{B}_n)$.

3.1 Real-valued processes with piecewise constant pdf

For our purposes, without loss of generality, it is convenient to focus on the non-empty subset of $\Omega(\mathbf{B}_n)$ of those processes having the random variables x_i uniformly distributed over each interval of the quantizing partition. By denoting with $\chi_I : \mathbb{R} \to \{0, 1\}$ the characteristic function of I, we define the special set of piecewise constant probability density functions (pdfs) as

$$D_{P_n} = \left\{ \phi = \sum_{i=0}^{2^n-1} a_i \chi_{I_i} : \|\phi\|_1 = 1, a_i \geq 0, I_i \in \mathcal{P}_n \right\}.$$

Accordingly, we denote with $\overline{\Omega}(\mathbf{B}_n) \subset \Omega(\mathbf{B}_n)$ the subset of real-valued stochastic processes such that for any $i \in \mathbb{N}$ the pdf of the random variable x_i belongs to D_{P_n}.

The fact that the set $\overline{\Omega}(\mathbf{B}_n)$ is not empty can be easily shown considering the following trivial stochastic process in $[0, 1)$. Given the process $\mathbf{B}_n = \{\mathbf{w}_i \in \{0, 1\}^n, i \in \mathbb{N}\}$, let us define $\{x_i \in [0, 1), i \in \mathbb{N}\}$ by construction:

if $\mathbf{w}_i = \mathbf{j}$ then uniformly pick at random $x_i \in I_j$.

3.2 Example

Starting from a given binary process $\mathbf{B}_2$, in this example we show a simple case in which it is possible to construct in an explicit closed form a process belonging

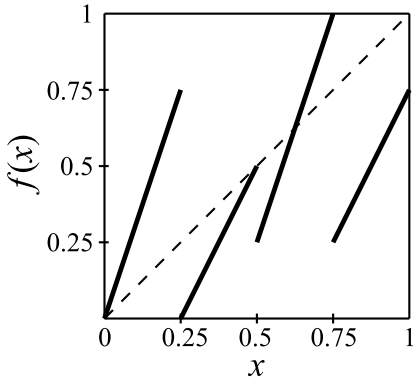


Figure 3: The source $\mathbf{B}_2$ in the example can be viewed as the result of a 2-bit quantization of the chaotic process $\mathbf{X}_f = \{f^i(x_0), i \in \mathbb{N}\}$, with x_0 being distributed according to (7). The stochastic process $\mathbf{X}_f$ is only one possible element of the infinite set $\overline{\Omega}(\mathbf{B}_2)$.

to $\overline{\Omega}(\mathbf{B}_2)$. We stress that this construction procedure *is not necessary for the design of the compression algorithm*, since as it is made clearer in the following Sections, it suffices to accept the abstract idea that a binary source $\mathbf{B}_n$ can be viewed as the result of a n-bit quantization of any real-valued stochastic process belonging to the $\overline{\Omega}(\mathbf{B}_n)$.

Let us consider the 4-states ergodic binary process $\mathbf{B}_2$ specified by both the Markovian transition probability matrix reported in Table 1 and the invariant probability mass vector $\mathbf{v}^* = (\frac{9}{29}, \frac{12}{29}, \frac{6}{29}, \frac{2}{29})$. We want to show a special element of $\overline{\Omega}(\mathbf{B}_2)$, that we denote with $\mathbf{X}_f$. Referring to the quantizing partition $I_0 = [0, 0.25)$, $I_1 = [0.25, 0.5)$, $I_2 = [0.5, 0.75)$, $I_3 = [0.75, 1)$, it can be proved that eq. (6) is satisfied if the two following conditions hold:
1) the statistical distribution for any x_i is described by the probability density function (pdf)

$$\begin{aligned}\phi_i(x) = \frac{4\cdot 9}{29}\chi_{I_0}(x) + \frac{4\cdot 12}{29}\chi_{I_1}(x) + \\ + \frac{4\cdot 6}{29}\chi_{I_2}(x) + \frac{4\cdot 2}{29}\chi_{I_3}(x);\end{aligned} \tag{7}$$

2) $\mathbf{X}_f$ has the following conditioned pdf

$$\phi(x_n|x_{n-1}) =$$
$$\begin{cases} \frac{4}{3}\chi_{I_0}(x_n) + \frac{4}{3}\chi_{I_1}(x_n) + \frac{4}{3}\chi_{I_2}(x_n), & \text{if } x_{n-1} \in I_0, \\ \frac{4}{2}\chi_{I_0}(x_n) + \frac{4}{2}\chi_{I_1}(x_n), & \text{if } x_{n-1} \in I_1, \\ \frac{4}{3}\chi_{I_1}(x_n) + \frac{4}{3}\chi_{I_2}(x_n) + \frac{4}{3}\chi_{I_3}(x_n), & \text{if } x_{n-1} \in I_2, \\ \frac{4}{2}\chi_{I_1}(x_n) + \frac{4}{2}\chi_{I_2}(x_n), & \text{if } x_{n-1} \in I_3. \end{cases}$$

Accordingly, it results that $P(\mathbf{w}_i = \mathbf{j}) = \int_{I_j} \phi_i(x)dx$ and $P(\mathbf{w}_{i+1} = \mathbf{j}_1|\mathbf{w}_i = \mathbf{j}_2) = \int_{I_{j_1}} \phi(x|x_i \in I_{j_2})dx$.

A hypothetical stochastic process satisfying the above conditions is the process $\mathbf{X}_f = \{f^i(x_0), i \in \mathbb{N}\}$ ruled by the chaotic map $f : [0, 1) \to [0, 1)$ shown in

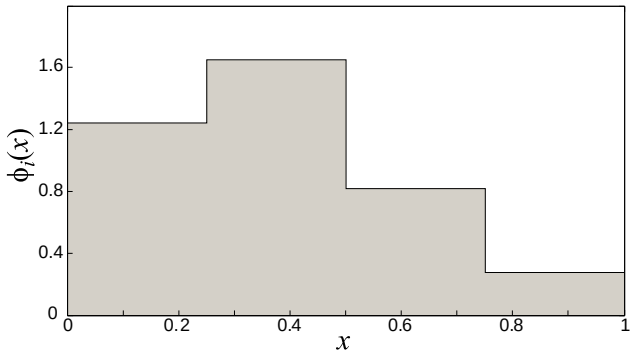


Figure 4: The stationary density (7) is an invariant density for the chaotic process $\mathbf{X}_f = \{f^i(x_0), i \in \mathbb{N}\}$.

Fig. 3, and having a random initial condition distributed according to the pdf (7). The pdf (7), depicted in Fig. 4, is invariant for this chaotic system and by construction it results

$$P(\mathbf{w}_0 = \mathbf{j}_0, \ldots \mathbf{w}_i = \mathbf{j}_i, \ldots) = \\ = P(x_0 \in I_{j_0}, \ldots, f^i(x_0) \in I_{j_i}, \ldots). \tag{8}$$

We omit the proof of this result, since further details about the link between Markovian processes and chaotic maps can be found in several books or papers [18].

4 Preliminaries: Chaotic mixing of stochastic processes

In this Section we report some theoretical results about the mixing properties of chaotic systems, adapted from [19, 20] and simplified to match the aim of this paper. Accordingly, we adopt the following notation and terminology. If I is an interval we denote with $\lambda(I)$ the Lebesgue measure of I. With reference to the Lebesgue integration theory, the notation $L^p(I)$ denotes the set of functions $f : I \to \mathbb{R}$ such that $\int_I |f(x)|^p dx < \infty$, with p positive integer, whereas $L^\infty(I)$ is the set of almost everywhere bounded measurable functions. We recall that $L^p(I)$ and $L^\infty(I)$ can be made Banach spaces with reference to the norms $\|f\|_p = (\int_I |f(x)|^p dx)^{\frac{1}{p}}$ and $\|f\|_\infty = \inf\{M \in \mathbb{R}^+ \text{ such that } \{x \in I : |f(x)| > M\} \text{ has zero measure}\}$, respectively.

4.1 Statistically stable mixing systems

Let $S : [0, 1) \to [0, 1)$ be a piecewise affine expanding (PWAE) map, i.e., the map is piecewise linear with $|dS/dx| > 1$, as in Fig. 3. We say that a pdf $\phi^* \in L^1$ is invariant for the map S if for any subset $A \subseteq [0, 1)$ it results

$$\int_A \phi^*(x)dx = \int_{S^{-1}(A)} \phi^*(x)dx,$$

which implies $P(S(x) \in A) = P(x \in A)$. By assuming $x_0 \in [0,1)$ as a random variable with pdf ϕ_0, let us focus on the sequence $\{S^p(x_0), p \in \mathbb{N}\}$. Even if S is deterministic $x_p = S^p(x_0)$ is a stochastic variable and we denote with ϕ_p its associated pdf. In this paper we refer to the following

Definition 2 *A PWAE map S is statistically stable iff $\forall \phi_0 \in D_{P_n}$ there exist an unique invariant pdf $\phi^* \in L^1([0,1))$ such that*

$$\lim_{p\to\infty} \|\phi_p - \phi^*\|_\infty = 0. \tag{9}$$

According to the above definition, for a statistically stable PWAE map as far as $p \to \infty$ the pdf of the random variable $S^p(x_0)$ becomes progressively uncorrelated with the distribution of the initial condition x_0, approaching an invariant (stationary) pdf ϕ^* that only depends on the map S. An efficient numerical method for analyzing the converging rate of the above limit can be found in [19, 21].

4.2 Mixing of $\overline{\Omega}(\mathbf{B}_n)$ stochastic processes

For a given ergodic binary source $\mathbf{B}_n$ let us now focus on a stochastic process $\mathbf{X} = \{x_i \in [0,1); i \in \mathbb{N}\}$ belonging to $\overline{\Omega}(\mathbf{B}_n)$, that is:

1. referring to the partition $\mathcal{P}_n$, the random variables x_i have constant pdf over each interval of $\mathcal{P}_n$;
2. the sequences of $\mathbf{B}_n$ are the n-bit quantization of the sequences of $\mathbf{X}$ through the partition $\mathcal{P}_n$.

If S is a statistically stable PWAE map, we can define the new stochastic process

$$\mathbf{Y}_p = \{y_i = S^p(x_i); i \in \mathbb{N}\},$$

for a given $p \in \mathbb{N}$. If the process $\mathbf{X}$ is ergodic, also $\mathbf{Y}$ is ergodic, since this is assured by the properties of PWAE maps [18, 20, 22]. It is worth noting that in general the process $\mathbf{Y}_p$ do not belong to $\overline{\Omega}(\mathbf{B}_n)$ since when coded through the partition $\mathcal{P}_n$ it may define a dynamics completely different from that one of $\mathbf{B}_n$. Moreover, even if the pdfs associated to the random variables y_i are piecewise constant, their discontinuity points may not fall on the endpoints of the partition $\mathcal{P}_n$, resulting to be not constant over each quantization interval [19, 21].

In this paper we have a special interest in those statistically stable PWAE maps having an uniform invariant pdf ϕ^*, as shown in the following example.

4.3 Example (Part I)

We introduce the Rényi chaotic map, that will be used in the next Sections as a reference example, to show how to design a lossy compressor based on a digitized version of this map. The Rényi map $T : [0,1) \to [0,1)$ is a PWAE map of the form $T(x) = \beta x \bmod 1$, with $\beta > 1$. In this paper we assume $\beta = 6$, i.e.,

$$T(x) = 6x \bmod 1. \tag{10}$$

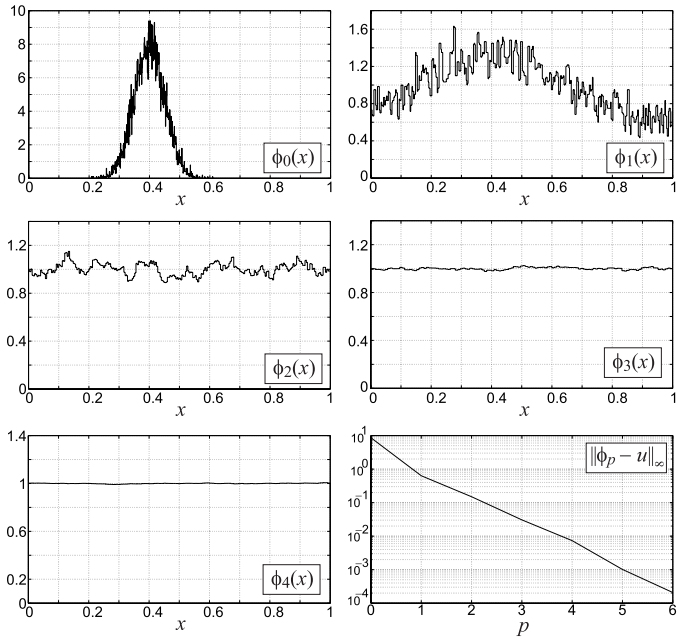


Figure 5: The pdfs associated to the random variables $y = T^p(x)$, for $p = 0, 1, \ldots, 4$. After 5 iterations the distance $\|\phi_5 - u\|_\infty$ is lower than 10^{-3}: the convergence is exponential, as confirmed in the log-scale plot of $\|\phi_p - u\|_\infty$. The pdfs were calculated exploiting a modified version of the accurate approach described in [19, 21].

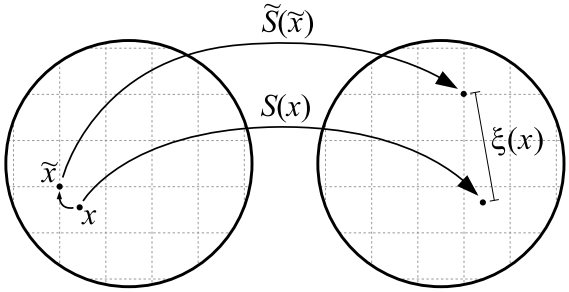


Figure 6: The effect of the finite-precision computation of a function S.

The above PWAE map is statistically stable, admitting the uniform pdf as the unique invariant pdf ϕ^* [17–20]. This means that, regardless of the pdf $\phi_0 \in D_{\mathcal{P}_n}$ associated to x, the greater is p and the more uniformly-distributed is the random variable $y = T^p(x)$, as stated in (9). In other words, $\forall \varepsilon > 0$ there exists p such that

$$\|\phi_p - u\|_\infty < \varepsilon, \tag{11}$$

where $u : [0,1) \to \{1\}$ is the uniform pdf. The number of iterations p necessary to satisfy the above inequality in general depends on ϕ_0 and ε, but *in cases of practical interest*, since the converging rate is exponential, few iterations suffices to obtain a reasonably accurate approximation of the uniform pdf u [17, 20]. Indeed, 5-6 iterations of the map T suffices to satisfy the inequality (11) with values of ε in the order of 10^{-4}, even for ϕ_0 quite different from the uniform pdf, e.g., see Fig. 5.

5 Main General Result: From chaos to pseudo-chaos, reliability analysis

Pseudo-chaos is obtained when a chaotic dynamical system is simulated using finite precision arithmetic algorithms. In detail, referring to a generic chaotic map $S : [0,1) \to [0,1)$, the pseudo-chaotic approximation of S is obtained in two steps. In the first step any point $x \in [0,1)$ is represented by a finite-precision point $\tilde{x}$ belonging to a finite set $\Lambda \subset [0,1)$, called the discrete domain. Accordingly we have the definition of the finite-precision point $\tilde{x} : [0,1) \to \Lambda$, as a function of x. In the second step the function S is approximated by a finite-precision approximating function $\tilde{S} : \Lambda \to \Lambda$ such that

$$\left| S(x) - \tilde{S}(\tilde{x}) \right| = \xi(x) < 1,$$

where the function ξ quantifies the quality of the approximation. Summarizing, the overall pseudo-chaotic approximation is defined by the composition $\tilde{S} \circ \tilde{x} : [0,1) \to \Lambda$.

If the function ξ assumes reasonably small values for all x, it is often said that $\tilde{S}$ 'shadows' S. Trying to relate the properties of systems S and $\tilde{S}$ can be a complex task, depending on the nonlinear functional forms of the two systems. For example, adopting a dynamical evolution point of view, Coomes et al. in [15]

proved some major results valid for hyperbolic diffeomorphisms, that represent a fundamental reference within the Shadowing Theory for chaotic systems.

In this paper we adopt a different point of view, since we are interested in relating the probability measures associated to S and $\tilde{S}$. In detail, for a given subset $I \subseteq [0,1)$, we are interested in the quantity

$$\mathcal{E}(I,\xi) = \left| P(S(x) \in I) - P(\tilde{S}(\tilde{x}) \in I) \right|.$$

For the aim of this paper, the set I is understood to be a sub-interval of $[0,1)$. As a general comment, it can be noted that regardless of $0 \leq \xi < 1$ the above quantity approaches 0 as $I \to [0,1)$. Indeed, as soon as the interval I covers the entire domain $[0,1)$, also the discrete domain $\Lambda \subset [0,1)$ is covered. Since S is measurable, the previous quantity can be written as

$$\mathcal{E}(I,\xi) = \left| P(x \in S^{-1}(I)) - P(\tilde{x} \in \tilde{S}^{-1}(I \cap \Lambda)) \right|.$$

Due to the effect of the approximation, in general it can happen that

$$\tilde{S}^{-1}(I \cap \Lambda) - (S^{-1}(I) \cap \Lambda) \neq \varnothing,$$

i.e., there are points mapped in I by the function $\tilde{S}$ that do not belong to $S^{-1}(I)$. About this issue, by defining

$$\alpha = \sup_{x \in S^{-1}(I)} \xi(x),$$

we have the following

Proposition 1 *Let $I \subseteq [0,1)$ be an interval with endpoints $a < b$, and let us assume that $\forall x \in S^{-1}(I)$ it results*

$$\left| S(x) - \tilde{S}(\tilde{x}) \right| < \alpha. \tag{12}$$

Accordingly, by denoting with ϕ the pdf of $S(x)$, if $b - a \geq 2\alpha$ then it results

$$\int_{a+\alpha}^{b-\alpha} \phi(x)dx \leq P\left(\tilde{S}(\tilde{x}) \in I\right) \leq \int_{a-\alpha}^{b+\alpha} \phi(x)dx. \tag{13}$$

Proof. See the Appendix I. □

It is worth noting that the above Proposition has a general validity, and that if the condition (12) holds in the whole interval $[0,1)$, then the result (13) holds for any interval $I \subseteq [0,1)$ with length greater than 2α. This is the case provided by the Shadowing Theory, that is valid for certain continuous maps [15]; in this paper we allow the chaotic map to be *not* continuous, as discussed in the next Sections.

5.1 Almost-uniform measure-preserving chaotic transformations

Under the hypotheses of Proposition 1, in this subsection we refine the theoretical result (13) assuming the random variable $S(x)$ almost uniformly-distributed, i.e., we assume

$$\|\phi - u\|_\infty \leq \varepsilon,$$

where $u : [0,1) \to \{1\}$ is the uniform pdf over $[0,1)$.

Accordingly, for any interval $I \subseteq [0,1)$ with endpoints $a < b$ it results

$$(b-a)(1-\varepsilon) \leq \int_I \phi(x)dx \leq (b-a)(1+\varepsilon). \tag{14}$$

If we now assume the interval I to have Lebesgue measure $\frac{1}{2^k}$, i.e., $b-a = \frac{1}{2^k}$, using (14) the inequality (13) can be rewritten as

$$\left(\frac{1}{2^k} - 2\alpha\right)(1-\varepsilon) \leq P\left(\tilde{S}(\tilde{x}) \in I\right) \leq \left(\frac{1}{2^k} + 2\alpha\right)(1+\varepsilon),$$

which implies

$$-\frac{\varepsilon}{2^k} - 2\alpha(1-\varepsilon) \leq P\left(\tilde{S}(\tilde{x}) \in I\right) - \frac{1}{2^k} \leq \\ \leq \frac{\varepsilon}{2^k} + 2\alpha(1+\varepsilon).$$

Focusing on the worst case, we finally have

$$\left|P\left(\tilde{S}(\tilde{x}) \in I\right) - \frac{1}{2^k}\right| \leq \frac{\varepsilon}{2^k} + 2\alpha(1+\varepsilon). \tag{15}$$

The similarity between (15) and (3) is clear: our idea is to use a discretized version of an almost-uniform measure-preserving chaotic transformation to boost the final entropy of a TRBG. In detail, by setting

$$\delta = \frac{\varepsilon}{2^k} + 2\alpha(1+\varepsilon)$$

such to satisfy any arbitrary worst-case level ASE_k, we have a precise constraint for the quantities ε and α. A compression method that exploits this idea will be discussed in the next Sections.

5.2 Example (Part II)

Let $S : [0,1) \to [0,1)$ be a chaotic map with $S(x)$ almost uniformly distributed. Furthermore, let $\tilde{S}$ be a pseudo-chaotic version of S, and let us define an ergodic binary source $\mathbf{B}_8$ by performing the symbolic coding of the sequence $\{\tilde{S}(\tilde{x}_i) \in [0,1); i \in \mathbb{N}\}$. In other words the binary source is defined by the rule

$$\mathbf{w} = \mathbf{j} \Leftrightarrow \tilde{S}(\tilde{x}) \in I_j,$$

where $\mathbf{j} = b_7 \ldots b_0$ is the binary representation of the integer $0 \leq j < 2^8$, and $\mathcal{P}_8 = \{I_0, \ldots, I_{255}\}$ is the partition made of $2^8 = 256$ equal intervals evenly dividing the domain $[0,1)$.

Let us focus on the Average Shannon Entropy of this source. According to the Proposition 1 and (5), in order to assure the ASE_8 to be greater than 0.95 bit/time-step, it suffices that for each interval of the partition $\mathcal{P}_8$ the probability for $\tilde{S}(\tilde{x})$ to belong to any interval I_i satisfies

$$\left|P(\tilde{S}(\tilde{x}) \in I_i) - \frac{1}{2^8}\right| < \delta \approx 2.79 \cdot 10^{-3}.$$

Accordingly, by setting

$$\delta = \frac{\varepsilon}{2^8} + 2\alpha(1+\varepsilon) \leq 2.79 \cdot 10^{-3}$$

we finally obtain

$$\alpha \leq \frac{0.71 - \varepsilon}{512(1+\varepsilon)}. \tag{16}$$

If ε is reasonably small the previous inequality can be approximated as

$$\alpha < \frac{0.71}{512}. \tag{17}$$

The determination of the quantity α mainly depends on the system $\tilde{S}$, as discussed in the next Section.

6 A TRBGs Lossy Compressor Based on the Discretized Rényi Map

In this Section we propose to use a pseudo-chaotic digitized version of the Rényi chaotic map (10) for the definition of a lossy compressor to process TRBGs. As discussed in the Introduction, even if we discuss a specific example based on the Rényi chaotic map, the results previously discussed have a general validity, and following the design approach discussed hereafter the general compression scheme of Fig. 1 can be obtained referring to any chaotic system belonging to the wide family of piecewise affine expanding maps having an almost uniform invariant measure. We have chosen the Rényi chaotic map since it is a not continuous map[3] and its pseudo-chaotic implementation involves very efficient hardware circuits, as it has been recently technically demonstrated even for high digitization accuracy (i.e., with n ranging up to 10^3) [16].

Assuming $n \in \mathbb{N}$, with $n > 1$, we define the set of dyadic rationals

$$\Lambda_n = \left\{ \frac{q}{2^n} \in \mathbb{N} : 0 \leq q < 2^n \right\},$$

and the discretized Rényi map $\tilde{T}_n : \Lambda_n \to \Lambda_n$ as

$$\begin{aligned} \tilde{T}_n\left(\frac{q}{2^n}\right) &= \frac{1}{2^n}\left(6q + \left\lfloor \frac{q}{2^{n-1}} \right\rfloor \mod 2^n\right) = \\ &= \frac{1}{2^n} \cdot \begin{cases} 6q \mod 2^n, & \text{if } 0 \leq q < 2^{n-1}, \\ 6q+1 \mod 2^n, & \text{if } 2^{n-1} \leq q < 2^n. \end{cases} \end{aligned} \tag{18}$$

The above digitized map belongs to a wide family of nonlinear permutations firstly investigated in [23]. For the above digitization of x it is assumed the truncation approximation strategy, i.e., if $x \in [0,1)$ it results

$$\tilde{x} = \frac{q(x)}{2^n} = \frac{\lfloor 2^n x \rfloor}{2^n}.$$

[3] This choice allows us to manage a case that can not be treated referring to the Shadowing Theory proposed by Coomes et al., that holds only for continuous maps.

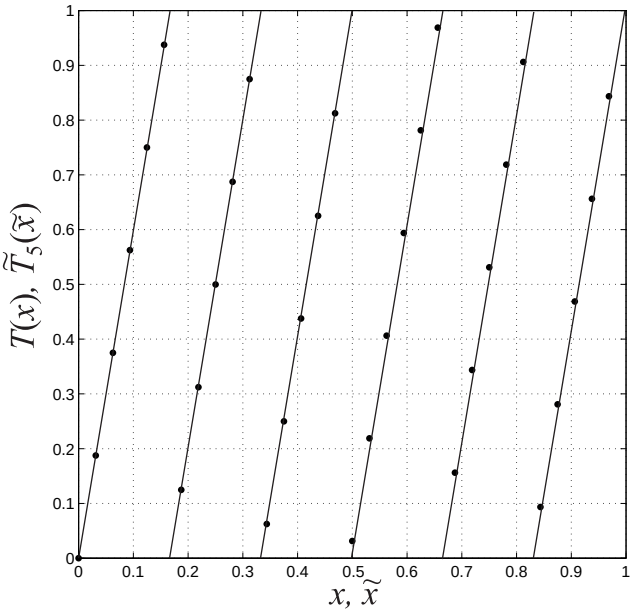


Figure 7: The Rényi chaotic map (10), superimposed to the digitized map (18), for $n = 5$. The digitized map is nonlinear, and the points lie on a slightly distorted lattice [23].

Accordingly, any point x can be written as

$$x = \tilde{x} + \xi(x) = \frac{q(x)}{2^n} + \xi(x),$$

where $\xi(x) \in [0, 2^{-n})$.

6.1 Properties of the Digitized Rényi Map

First of all, we stress that (18) is a nonlinear permutation of Λ_n, and any iterated $\tilde{T}_n^j$ is a bijection in Λ_n, as discussed in [23]. Moreover, the digital architecture implementing (18) is a n-bit machine, and for this reason the parameter n coincides with the digital resolution of the pseudo-chaotic system [16, 23]. Let us now show now that the the discretized Rényi map approximates the Rényi map in the sense specified by the following

Theorem 2 *Let us consider the interval*

$$K = \left[\frac{6^{p+1}}{5 \cdot 2^n}, 1 - \frac{6^{p+1}}{5 \cdot 2^n} \right),$$

with $p, n \in \mathbb{N}$ *and* $n > (p+1)\log_2 6 - 2$. *For any* $x \in T^{-p}(K)$ *it results*

$$\left| T^p(x) - \tilde{T}_n^p(\tilde{x}) \right| < \frac{6^{p+1}}{5 \cdot 2^n}. \tag{19}$$

Proof. See the Appendix II. □

According to the previous Theorem, for any given p a number n can be set such that the interval K covers $[0, 1)$ almost entirely, with

$$\left| T^p(x) - \tilde{T}_n^p(\tilde{x}) \right| < \frac{6^{p+1}}{5 \cdot 2^n} = \alpha, \tag{20}$$

α being arbitrarily small.

Theorem 2 does not tell us what may happen outside the interval K, at the borders of the phase space $[0, 1)$: if $x \notin T^{-p}(K)$ the 'shadowing error' made by the digitized map $\tilde{T}_n$ after p iterations can be much greater than the upper bound (19), due to the effects of the map discontinuities. As discussed in the following example, in practical cases this fact has negligible consequences on the working of the proposed compression method, since the dynamics of the digitized system is typically observed adopting a coarse-grained resolution, greater than α.

6.2 Example (Part III)

Let us focus on the binary source $\mathbf{B}_8$ defined in the previous example of Subsection 5.2. We have stated that in order to have an ASE_8 greater than 0.95 bit/time-step for any interval I_i of the partition $\mathcal{P}_8$ it must result $|P(\tilde{S}(\tilde{x}) \in I_i) - 2^{-8}| < \delta \approx 2.79 \cdot 10^{-3}$, and we have also seen that from (17) this latter result is assured if for all $x \in S^{-1}(I_i)$ we have

$$|S(x) - \tilde{S}(\tilde{x})| < \alpha < \frac{0.71}{512}. \tag{21}$$

Is it possible to obtain such result for a not-continuous system like the Rényi map T? Since (20) does not hold for all $x \in [0, 1)$, it seems not. Actually, in this example we will show that the effects of the discontinuities for maps like the Rényi map are typically negligible.

Accordingly, we assume the map S to be the p-th iterate of the Rényi map T, i.e., $S = T^p : [0, 1) \to [0, 1)$. As discussed before, in most cases of practical interest a small number of iterations ($p \approx 5$) assures that the random variable $S(x)$ is almost uniformly distributed. Moreover, we assume the discrete function $\tilde{S}$ to be the pseudo-chaotic version of the p-times iterated Rényi map, i.e. $\tilde{S} = \tilde{T}_n^p$. If the random variable $S^p(x)$ is almost uniformly distributed (such to neglect the term ε in (16)), combining (19) with (21) we obtain

$$\frac{6^{p+1}}{5 \cdot 2^n} < \frac{0.71}{512},$$

which yields

$$n > n_{\min} = \left\lceil p \log_2 6 + 9 + \log_2 \frac{6}{0.71 \cdot 5} \right\rceil. \tag{22}$$

The result of (22), for different values of p, is reported in the first two rows of Table 2. The third row of this Table discloses the concept of compression developed in this paper: in order to obtain a binary source $\mathbf{B}_8$ generating 8-bit words with a worst-case ASE_8 entropy greater than 0.95 bit/time-step, we operate on the n-bit state of a p-times iterated pseudo-chaotic machine, with $n > n_{\min} > 8$. We will discuss this aspect in detail in the next Sections.

Table 2: Minimum number of bits satisfying (22) and best obtainable compression ratio, for different values of p

p	1	2	3	4	5	6	7	8
$n_{\min}$	$\lceil 12.34 \rceil$	$\lceil 14.92 \rceil$	$\lceil 17.51 \rceil$	$\lceil 20.10 \rceil$	$\lceil 22.68 \rceil$	$\lceil 25.27 \rceil$	$\lceil 27.85 \rceil$	$\lceil 30.44 \rceil$
Best Compression Ratio	1:1.625	1:1.875	1:2.25	1:2.625	1:2.875	1:3.25	1:3.5	1:3.875

Since we are dealing with an example, let us assume $p = 5$ and $n = 23$, i.e., we are assuming that in most practical cases $S(x)$ is almost uniformly distributed. In such case, according to (20), $\alpha \approx 1.1 \cdot 10^{-3}$, and referring to Theorem (2) we have that

$$K = [0.0011, 0.9989),$$

and for all $x \in T^{-5}(K)$, (21) is verified. Summarizing, we can state that for any interval I_i of the partition $\mathcal{P}_8$ that lyes inside K, it results

$$\left| P(\tilde{S}(\tilde{x}) \in I_i) - \frac{1}{2^8} \right| < \delta \approx 2.79 \cdot 10^{-3}. \tag{23}$$

What happens in those intervals of $\mathcal{P}_8$ that do not lye inside K? Since the length of each of these intervals is $\frac{1}{2^8}$, the only two intervals of $\mathcal{P}_8$ not completely inside K are $I_0 = [0, \frac{1}{2^8})$ and $I_{255} = [\frac{255}{2^8}, 1)$. Even if we assume the worst-case $P(\tilde{S}(\tilde{x}) \in I_0 \cup I_{255}) = 0$, the summation (2) is reduced by only two terms. As a result, the ASE_8 is scarcely influenced by this change, since the two terms in (2) approximately weigh $\approx \frac{2 \cdot ASE_8|_{wc}}{2^8}$. Actually, we stress that this scenario is extremely pessimistic, since the two intervals I_0 and I_{255} mostly overlap with K. Indeed, focusing on I_0, we have $J = I_0 \cap K = [\alpha, \frac{1}{2^8})$, with length $\lambda(J) \approx \frac{1}{256} - 1.1 \cdot 10^{-3} \approx 2.8 \cdot 10^{-3} \approx 2.55\alpha$. Using (13) and (14) we have

$$0.55\alpha(1 - \varepsilon) \leq P(\tilde{S}(\tilde{x}) \in I_0 \cap K) \leq 4.55\alpha(1 + \varepsilon),$$

that implies that $P(\tilde{S}(\tilde{x}) \in I_0)$ is at least about $0.55\alpha > 0$, in the worst case.

In general, adopting an heuristic point of view, we can analyze the statistical effects of the discontinuities of the map at the borders of $[0, 1)$, considering that $S(x)$ is almost uniformly distributed. In such situation, since $\tilde{T}_n$ is a permutation of Λ, and due to the uniformity over $[0, 1)$ of the digitization strategy $\tilde{x}(x)$, it is reasonably to expect that $P(\tilde{S}(\tilde{x}) \in I_0) \approx P(\tilde{S}(\tilde{x}) \in I_{255})$ and, from the normalization property

$$\sum_{i=0}^{255} P(\tilde{S}(\tilde{x}) \in I_i) = 1,$$

that also for I_0 and I_{255} the ineq. (23) is statistically satisfied, approximately.

7 Definition of the Lossy Compressor

Until now, once set the $\mathrm{ASE}_k|_{wc}$ of the target binary source $\mathbf{B}_k$, we have explained, referring to the Rényi map, how to design the digital resolution n of the digitized map $\tilde{T}_n$ (i.e., how to set α), assuming the random variable $T^p(x)$

The Compression Algorithm

1. Collect n bits $b_{n-1}, \ldots, b_0$ to form $\tilde{x} = \sum_{i=0}^{n-1} b_i 2^{i-n}$;
2. Compute $\tilde{y} = \tilde{T}_n^p(\tilde{x})$;
3. Output the k most significant bits of $\tilde{y}$, with $k \leq n$.

Figure 8: The simple working algorithm of the proposed compressor.

almost uniformly distributed. Since our aim is the design of a lossy compressor to be applied to TRBGs, we still have to describe the overall compression algorithm, determining the link between the output of the TRBG and the random variable x, which is first digitized to get $\tilde{x}$ and then processed to obtain $\tilde{T}_n(\tilde{x})$.

Before discussing these theoretical aspects, related to the design of the compressor parameters, we first present the compression method, that is very simple, by means of Figs. 1 and 8. Even if we focus on the Rényi map, we stress that the presented method can be generalized, referring to any pseudo-chaotic system $\tilde{S}$ that is linked to the original map S by the eq. (12), and such that S is a statistically stable map having the uniform pdf as invariant density ϕ^*.

Referring to the Fig. 8, the first step of the compression algorithm is to collect n bits $b_{n-1}, \ldots, b_0$ from the TRBG output, so to define the dyadic rational $\tilde{x} = \frac{m}{2^n} = \sum_{i=0}^{n-1} b_i 2^{i-n}$. It is worth noting that, referring to the theoretical discussion presented in the previous Sections, the n bits collected from the TRBG define a discrete variable $\tilde{x}$ that can be interpreted as the fixed point representation of the n-bit quantization of x by means of the partition $\mathcal{P}_n$. This point of view reveals the link between the output of the TRBG and the random variable x, that according the compressor theoretical explanation is first used to get $\tilde{x}$ and then processed to obtain $\tilde{T}_n(\tilde{x})$.

Since the condition (6) is the only constraint that relates $\mathbf{B}_n$ with the hypothetical process $\mathbf{x}$, we are free to assume the hypothetical process $\mathbf{x}$ belonging to $\overline{\Omega}(\mathbf{B}_n)$. Accordingly, the piecewise constant stationary probability density function ϕ_0 associated to the random variable x is such that, for any $I_j \in \mathcal{P}_n$,

$$\phi_0(x) = 2^n \sum_{j=0}^{2^n-1} P(\mathbf{w} = \mathbf{j}) \chi_{I_j}(x).$$

As a result, the levels of the pdf ϕ_0 associated to x are directly proportional to generation probabilities associated to the n-bit words issued by the binary source $\mathbf{B}_n$, and the chaotic map S has to be chosen such that for any ϕ_0 of practical interest the random variable $S(x)$ is almost uniformly distributed. In the example discussed in the previous Section we chose $S = T^p$, where T is the Rényi chaotic map and $p = 5$. When dealing with a different PWAE map, the appropriate number of iterations can be found following the same reasoning discussed in the Subsection 4.3, using for the accurate calculations the same theoretical approach discussed in [19, 21].

As a final step of the compression algorithm, the overall output is simply obtained selecting the k most significant bits of $\tilde{T}^p(\tilde{x})$, in accordance with the theory developed from Section 2 to Section 6, that assures the final binary source $\mathbf{B}_k$ to have an ASE_k greater than a given $\mathrm{ASE}_k|_{wc}$. It is worth noting that outputting the k most significant bits of $\tilde{T}_n(\tilde{x})$ agrees with performing k-bit

quantization of $\tilde{T}_n(\tilde{x})$ by means of the partition $\mathcal{P}_k$.

7.1 Some remarks on the implementation of the compressor

We spend only few comments on the easy digital implementation of the compressor, since the overall device consists in a digital block computing $\tilde{T}_n^p(\tilde{x})$ and outputting the $k < n$ most significant bits of the result. If we denote with X_n the binary vector representing $\tilde{x}$, from an algorithmic point of view the calculation of $\tilde{T}_n(\tilde{x})$ in modular arithmetic corresponds to adding $4X_n + 2X_n$, taking the n less significant bits of the results and forcing its least significant bit to be equal to the most significant bit of X_n. Indeed, referring to the definition (18), we have that $m = 2^n\tilde{x} > 2^{n-1}$ if and only if the most significant bit of X_n is '1'. Further considerations about the generic implementation of the Rényi pseudo-chaotic map can be found in [23], whereas when the digital resolution n is very high (i.e., in the orders of $10^2 \div 10^3$), the reader can refer to [16].

7.2 Application to an actual TRBG circuit

In this Subsection we apply the proposed compression method to the output of the well known chaotic TRBG proposed in [19, 24, 25]. The authors could implement a prototype of this TRBG on a Field Programmable Analog Array, as discussed in [25]. The statistical characteristics of the TRBG depend on the parameters of the chaotic map, implemented with analog hardware and therefore strongly susceptible to the variability of the circuit parameters, e.g., due to the tolerances of the fabrication process or to some changes in the environmental working conditions. Despite those methods proposed in literature to counteract this problem [25], the ASE of the TRBG is always not greater than a maximum theoretical limit that depends on a number of factors, among which there is the electronic noise [26]. Accordingly, in order to assure and adequate level of entropy for the TRBG, compression of the generated sequences have to be performed. Since we were interested in testing the suitability of the proposed compression method, in this Section we set the entropy of the considered TRBG much lower than its typical values, forcing the analog circuits to work in an *extremely worst case* condition.

Taking as an observing variable the distribution of generated 8-bit words, in Fig. 9 we reported the effect of the proposed compression method for different compression ratios, applying the method to an acquired sequence of 10^6 random bits. In the upper-left plot the TRBG generation probabilities for the 256 possible words of 8 bits is shown (no compression applied). As it can be seen by this first plot, most of the words are never generated, determining an ASE_8 as low as ≈ 0.81 bit/time-step, as reported in the lower right plot (for $p = 0$). Adopting the design criteria discussed in the example of Table 2, we then applied to the sequence an increasing level of compression, taking $n = 13, 15, 18, 21, 23, 26, 31$ bits at a time for the definition of $\tilde{x}$, calculating respectively $p = 1, 2, 3, 4, 5, 6, 7, 8$ iterations of the digitized Rényi chaotic map $\tilde{T}_n(\tilde{x})$, and outputting the 8 most significant bits of the result. Accordingly, the correspondent compression ratios agree with those values reported in the third row of the Table 2.

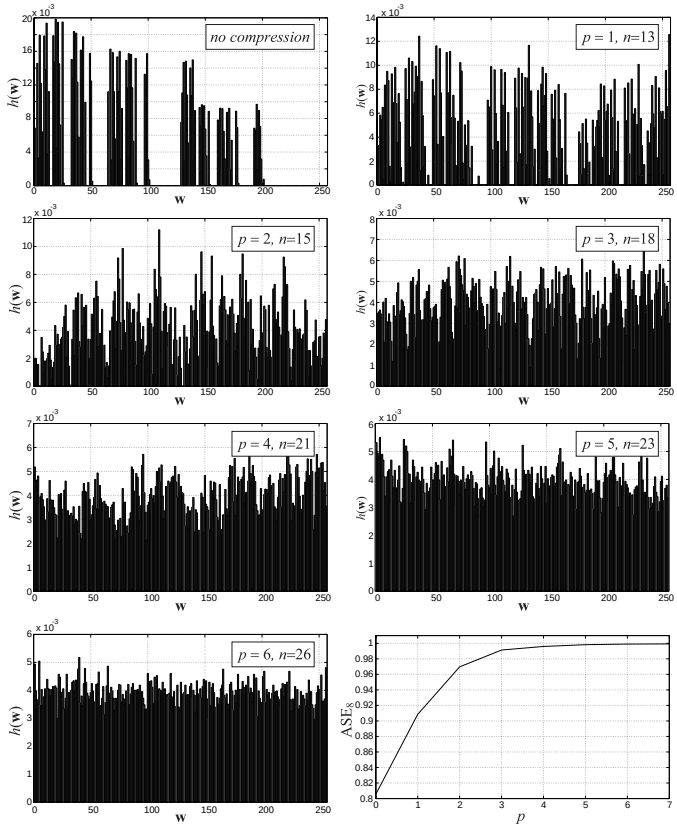


Figure 9: Effect of the proposed compression method on the output of an actual TRBG characterized by an $\mathrm{ASE}_8 \approx 0.81$ bit/time-step.

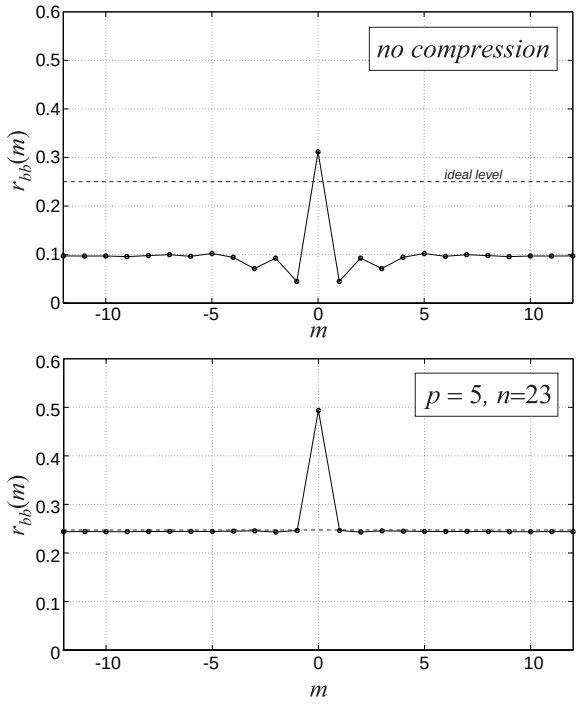


Figure 10: Effect of the proposed compression method on both the bias and the autocorrelation between generated bits.

The experiment well confirmed the suitability of the proposed compression method, and the results deserve some further comments. From the lower-right plot of Fig. 9, it can be noticed that the estimated ASE_8 for $p = 5$ is greater than 0.99 bit/time-step, that is much greater than the design worst case $\mathrm{ASE}_8|_{wc} = 0.95$ bit/time-step. This is normal, since all the theoretical results discussed in this paper are based on worst-case sufficient conditions. This is the reason why we defined the proposed method a *deep lossy compressor*, as discussed in the Introduction: in most cases, according to the design criteria discussed in this paper, the TRBG is over-compressed, and some information is lost. On the other hand, it is worth stressing that when dealing with Random Number Generators, especially in cryptography, the good statistical quality of generated sequence is the primary target requirement to be satisfied, even at the cost of a limited throughput reduction. The effect of compression is beneficial even for the correction of both the bias and the autocorrelation between generated bits, as shown in Fig. 10. In this figure, if $b_0, b_1, \ldots$ is the generated binary sequence, the autocorrelation $r_{bb}(m) = E\{b_i b_{i+m}\}$ is depicted for m ranging up to 12, for the original TRBG and after compression (for $m > 12$ the function is almost constant in both cases). Since $b_i \in \{0, 1\}$, the value $r_{bb}(0)$ corresponds to the mean value of the sequence, which should be 0.5 for an ideal TRBG. As it can be seen, both the bias and the autocorrelation between generated bits are relevant for the not-compressed source, whereas after the compression the levels are almost restored to their ideal values ($r_{bb}(0) = 0.5$ and $r_{bb}(m) = 0.25$, for $m \neq 0$).

For the sake of completeness, we also evaluated the compressed sequences by means of the severe NIST SP800.22 standard statistical tests [3]. As it was expected, for a given TRBG the statistical quality of the compressed sequences increases with the compression rate. For example, for the extremely poor TRBG evaluated in this Section, a compressor with parameters $(k, n, p) = (8, 8, 31)$ suffices to pass all of the NIST tests, when evaluating sequences of 10^6 bits. On the other hand, as far as the compression rate is reduced, several tests in the NIST standard were progressively failed. By letting the analog circuits of the same TRBG device work under normal operative conditions[4] a less complex compressor with parameters $(k, n, p) = (8, 5, 23)$ sufficed to pass all the NIST tests, with a compression ratio equal to $1 : 2.875$. The statistical tests were applied to 100 output sequences of 10^6 bits [3].

Even if a strong compression can turn a poor TRBG in a cryptographically good device, it is worth recalling that any compression algorithm can not protect against an hardware failure of the TRBG, since compression can only assure the output ASE to be not lower than the input ASE. Accordingly, other countermeasures have to be taken into account when dealing with cryptographic applications, e.g., by performing a bit-by-bit XOR-operation of the compressed sequence with a sequence generated by a cryptographically strong Pseudo Random Bit Generator (PRBG) [27].

7.3 Comparison with the direct under-sampling method.

In the direct under-sampling method the output bits are obtained by decimating the input sequence, or, more in general, a group of k bits is output for each group of n input bits, with $k < n$. In order to highlight the different working of the proposed compression method with the direct under-sampling approaches, we stress that in our case, before outputting $k < n$ bits, we build an advantageous coding of the sequence that boosts the final ASE, as theoretically demonstrated in the previous Sections. It can be easily shown that for binary processes with vanishing statistical dependencies the direct under-sampling method can boost the ASE up to the asymptotic value of $-\mu \log_2 \mu - (1 - \mu) \log_2 \mu$, that is the entropy of a Bernoulli binary source with alphabet $\{0, 1\}$ and mean value μ. For the poor TRBG previously analyzed, $\mu \approx 0.37$ and the ASE asymptotic value is $\approx$0.95 bit/time-step, referring to a compression rate that approaches zero in the limit case (i.e., $k = 1$ and $n \to \infty$).

8 Conclusions

It this paper we have proposed a compression method that exploits pseudo-chaotic systems, to be applied to True Random Bit Generators (TRBGs). The theory used to explain the working and the design of the compression scheme required to project some results achieved within the Ergodic Theory for chaotic systems on the world of digital pseudo-chaos. To this aim, we have proposed a weaker and more general interpretation of the Shadowing Theory, focusing on probability measures, rather than on single chaotic trajectories. We have proved

[4] I.e., for the analog circuits we removed the forced worst case operating conditions that determined the bad TRBG device with the 8-bit word distribution depicted in upper-left plot of Fig. 9.

theoretically how to design the overall compression scheme, in order to assure the final entropy of the compressed TRBG to be arbitrarily close to the maximum theoretical limit of 1 bit/time-step. Even if we discussed a specific example based on the Rényi chaotic map, the results proved in this paper are general, and adopting the proposed design approach the general compression scheme of Fig. 1 can be adapted to any chaotic system belonging to the wide family of piecewise affine expanding maps and having an almost uniform invariant pdf. The reliability of the proposed approach has been shown applying the compression scheme to an actual True Random Bit Generator prototype circuit.

9 APPENDIX - Proof of the Proposition 1

Proposition 1: Let $I \subseteq [0,1)$ be an interval with endpoints $a < b$, and let us assume that $\forall x \in S^{-1}(I)$ it results

$$\left|S(x) - \tilde{S}(\tilde{x})\right| < \alpha = \sup_{x \in S^{-1}(I)} \xi(x). \tag{24}$$

Accordingly, by denoting with ϕ the pdf of $S(x)$, if $b - a \geq 2\alpha$ then it results

$$\int_{a+\alpha}^{b-\alpha} \phi(x)dx \leq P\left(\tilde{S}(\tilde{x}) \in I\right) \leq \int_{a-\alpha}^{b+\alpha} \phi(x)dx.$$

Proof. From (24) it immediately results that there exists the subinterval $I' = (a+\alpha, b-\alpha) \subset I$ such that if $S(x) \in I'$ then $\tilde{S}(\tilde{x}) \in I$. Indeed, $S^{-1}(I') \subset S^{-1}(I)$ and if $x \in S^{-1}(I')$ then $\tilde{S}(\tilde{x}) \in I$, due to $|S(x) - \tilde{S}(\tilde{x})| < \alpha$ and $S(x) \in I'$. Accordingly, the probability for $\tilde{S}(\tilde{x})$ to belong to I must be not lower than $P(S(x) \in I')$, i.e.,

$$\int_{a+\alpha}^{a-\alpha} \phi(x)dx \leq P(\tilde{S}(\tilde{x}) \in I).$$

On the other hand, following the same reasoning, from (24) it follows that there exists the interval $I'' = (a-\alpha, b+\alpha) \supset I$ such that in the worst case all the points $x \in S^{-1}(I'') \supset S^{-1}(I)$ trigger approximated calculations $\tilde{S}(\tilde{x})$ that fall into I, such that if $S(x) \in I''$ then $\tilde{S}(\tilde{x}) \in I$. Accordingly, it must necessarily result

$$P(\tilde{S}(\tilde{x}) \in I) \leq \int_{a-\alpha}^{b+\alpha} \phi_1(x)dx,$$

concluding the proof. □

10 APPENDIX - Proof of Theorem 2

We first need the following

Lemma 1 *Let $\frac{q_0}{2^n} \in \Lambda_n$ and $z > \frac{q_0}{2^n}$. If the restriction of the Rényi map T to the interval $J = \left[\frac{q_0}{2^n}, z\right)$ is continuous, for any $x \in J$ and for any $m \in \mathbb{N}$, with $q_0 \leq m \leq \lfloor 2^n z \rfloor$, it results*

$$\left|T(x) - \tilde{T}_n\left(\frac{m}{2^n}\right)\right| < 6\left|x - \frac{m}{2^n}\right| + \frac{1}{2^n}.$$

Proof. We note that the restriction of T to the interval J is linear, and that for any $x_1, x_2 \in J$ we have $|T(x_2)-T(x_1)| = 6|x_2-x_1|$. Moreover, we note that for any $\frac{m}{2^n} \in \Lambda_n$ it results $\left|T\left(\frac{m}{2^n}\right) - \tilde{T}_n\left(\frac{m}{2^n}\right)\right| < \frac{1}{2^n}$. Accordingly, we have

$$\begin{aligned}
&\left|T(x) - \tilde{T}_n\left(\frac{m}{2^n}\right)\right| = \\
&= \left|T(x) - T\left(\frac{m}{2^n}\right) + T\left(\frac{m}{2^n}\right) - \tilde{T}_n\left(\frac{m}{2^n}\right)\right| \leq \\
&\leq \left|T(x) - T\left(\frac{m}{2^n}\right)\right| + \left|T\left(\frac{m}{2^n}\right) - \tilde{T}_n\left(\frac{m}{2^n}\right)\right| < \\
&< 6\left|x - \frac{m}{2^n}\right| + \frac{1}{2^n},
\end{aligned}$$

concluding the proof. □

We can now prove Theorem 2, i.e., the discretized Rényi map approximates the Rényi map in the sense specified by the following

Theorem 2: Let us consider the interval

$$K = \left[\frac{6^{p+1}}{5 \cdot 2^n}, 1 - \frac{6^{p+1}}{5 \cdot 2^n}\right),$$

with $p, n \in \mathbb{N}$ and $n > (p+1)\log_2 6 - 2$. For any $x \in T^{-p}(K)$ it results

$$\left|T^p(x) - \tilde{T}_n^p(\tilde{x})\right| < \frac{6^{p+1}}{5 \cdot 2^n}. \tag{25}$$

Proof. Let us assume the Rényi map continuous on intervals containing the points $T^j(x)$ and $\tilde{T}_n^j(\tilde{x})$, for $j = 0, 1, \ldots, p-1$. From the Lemma 1 we have

$$\left|T^{j+1}(x) - \tilde{T}_n^{j+1}(\tilde{x})\right| < 6\left|T^j(x) - \tilde{T}_n^j(\tilde{x})\right| + \frac{1}{2^n},$$

and proceeding by induction it is easy to prove that for $j > 0$

$$\begin{aligned}
\left|T^j(x) - \tilde{T}_n^j(\tilde{x})\right| &< 6^j|x-\tilde{x}| + \frac{1}{2^n}\sum_{r=0}^{j-1} 6^r = \\
&= 6^j|x-\tilde{x}| + \frac{6^j - 1}{5 \cdot 2^n} < \frac{6^{j+1}}{5 \cdot 2^n}.
\end{aligned} \tag{26}$$

Since the above inequality agrees with (25) by setting $j = p$, the proof is completed if we show that if $x \in T^{-p}(K)$ then some intervals exist containing the points $T^j(x)$ and $\tilde{T}_n^j(\tilde{x})$, for $j = 0, 1, \ldots, p-1$, and in which the Rényi map is continuous. To this aim, we start proving that any point $x \in T^{-j}(K)$ is at least $\frac{6^{p+1-j}}{5 \cdot 2^n}$ far away from any of the Rényi map discontinuity points. Indeed, if we denote with t an arbitrary discontinuity point of T, the set $T^{-j}([0, \frac{6^{p+1}}{5 \cdot 2^n}))$ contains intervals of the form $[t, t + \frac{6^{p+1-j}}{5 \cdot 2^n})$, that can not contain points of $T^{-j}(K)$. A same reasoning holds when considering the set $T^{-j}([1 - \frac{6^{p+1}}{5 \cdot 2^n}, 1))$. Accordingly, if $x \in T^{-p}(K)$ then for $j = 0, \ldots, p-1$ any point $T^j(x)$ is far away at least $\frac{6^{j+1}}{5 \cdot 2^n}$ from the discontinuity points, i.e.,

$$|T^j(x) - t| > \frac{6^{j+1}}{5 \cdot 2^n}.$$

On the basis of (26), the points $\tilde{T}_n^j(\tilde{x})$ are always between $T^j(x)$ and the discontinuity points, i.e., there exists an interval containing both of the points and such that the Rényi map is continuous. □

References

[1] D. Salomon, *Data Compression: The Complete Reference*. Springer-Verlag, 2007.

[2] P. L'Ecuyer and R. Proulx, "About polynomial-time "unpredictable" generators," in *Proceedings of the 1989 Winter Simulation Conference*. IEEE Press, Dec 1989, pp. 467–476.

[3] "NIST Special Publication 800-22 Rev.1a: A statistical test suite for random and pseudorandom number generators for cryptographic applications," Apr. 2010.

[4] L. Blum, M. Blum, and M. Shub, "A simple unpredictable pseudo-random number generator," *SIAM Journal on Computing*, vol. 15, no. 2, pp. 364–383, 1986.

[5] 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*, ser. Lecture Notes in Computer Science, C. Koç, D. Naccache, and C. Paar, Eds. Springer-Verlag, 2002, no. 106, pp. 431–449.

[6] 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, pp. 109–119, 2007.

[7] R. Barak, B. amd Shaltiel and E. Tromer, "True random number generators secure in a changing environment," in *Proc. CHES 2003, LNCS 2779*. Springer, 2003, pp. 166–180.

[8] L. Ziv, "The universal lz77 compression algorithm is essentially optimal for individual finite-length n-blocks," *IEEE Trans. Information Theory*, vol. 55, no. 5, pp. 1941–1944, May 2009.

[9] S. Hwang and C. Wu, "Unified VLSI systolic array design for LZ data compression," *IEEE Trans. VLSI Syst.*, vol. 9, no. 4, Ago. 2001.

[10] H. Hashempour1 and F. Lombardi, "Compression of vlsi test data by arithmetic coding," in *Proceedings of the 19th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems*, 2004.

[11] N. Ranganathan and S. Henriques, "High-speed VLSI designs for Lempel-Ziv-based data compression," *IEEE Trans. Circ. Sys. II*, vol. 40, no. 2, Feb 1993.

[12] J. Von Neumann, "Various techniques used in connection with random digits," *Applied Mathematics Series, U.S. National Bureau of Standards*, vol. 12, pp. 36–38, 1951.

[13] P. Elias, "The efficient construction of an unbiased random sequence," *Annals of Mathematical Statistics*, vol. 43, no. 3, pp. 865–870, 1972.

[14] A. Juels, M. Jakobsson, E. Shriver, and B. K. Hillyer, "How to turn loaded dice into fair coins," *IEEE Transaction on Information Theory*, vol. 46, no. 3, pp. 911–352, 2000.

[15] B. Coomes, H. Kocak, and K. Palmer, *Six Lectures on Dynamical Systems*. World Scientific, 1996, ch. Shadowing in Discrete Dynamical Systems, pp. 163–211.

[16] T. Addabbo, D. De Caro, A. Fort, N. Petra, S. Rocchi, and V. Vignoli, "Efficient implementation of pseudochaotic piecewise linear maps with high digitization accuracies," *International Journal of Circuit Theory and Applications*, 2010.

[17] T. Addabbo, M. Alioto, A. Fort, S. Rocchi, and V. Vignoli, "A variability-tolerant feedback technique for throughput maximization of TRBGs with predefined entropy," *Journal of Circuits, Systems and Computers*, vol. 19, no. 4, pp. 1–17, 2010.

[18] A. Lasota and M. C. Mackey, *Chaos, Fractals and Noise - Stochastic Aspects of Dynamics*, 2nd ed. Springer, 1994.

[19] T. Addabbo, A. Fort, S. Rocchi, D. Papini, and V. Vignoli, "Invariant measures of tunable chaotic sources: Robustness analysis and efficient computation," *IEEE Transactions on Circuits and Systems - I*, vol. 56, no. 4, pp. 806–819, 2009.

[20] A. Boyarsky and P. Góra, *Laws of Chaos*. Birkhäuser, 1997.

[21] T. Addabbo, A. Fort, D. Papini, S. Rocchi, and V. Vignoli, "An efficient and accurate method for the estimation of entropy and other dynamical invariants for piecewise affine choatic maps," *International Journal of Bifurcation and Chaos*, vol. 19, no. 12, pp. 4175 – 4195, 2009, accepted.

[22] P. Walters, *An Introduction to Ergodic Theory*. Springer, 1982.

[23] T. Addabbo, M. Alioto, A. Fort, A. Pasini, S. Rocchi, and V. Vignoli, "A class of maximum-period nonlinear congruential generators derived from the Rényi chaotic map," *IEEE Transactions on Circuits and Systems - I*, vol. 54, no. 4, pp. 816–828, 2007.

[24] T. Stojanovski and L. Kocarev, "Chaos-based random number generator – part I: Analysis," *IEEE Transactions on Circuits and Systems I*, vol. 48, no. 3, pp. 281–288, 2001.

[25] T. Addabbo, M. Alioto, A. Fort, S. Rocchi, and V. Vignoli, "A feedback strategy to improve the entropy of a chaos-based random bit generator," *IEEE Transaction on Circuits and Systems – part I*, vol. 53, no. 2, pp. 326–337, 2006.

[26] T. Addabbo, A. Fort, S. Rocchi, and V. Vignoli, "Exploiting chaotic dynamics for a-d converter testing," *International Journal of Bifurcation and Chaos*, vol. 20, no. 4, pp. 1099–1118, 2010.

[27] A. Menezes, P. Van Oorschot, and S. Vanstone, *Handbook of Applied Cryptography*. CRC Press, 1997.