

Chaos-Based Random Number Generators—Part I: Analysis

Toni Stojanovski and Ljupčo Kocarev, *Senior Member, IEEE*

Abstract—This paper and its companion (Part II) are devoted to the analysis of the application of a chaotic piecewise-linear one-dimensional (PL1D) map as random number generator (RNG). Piecewise linearity of the map enables us to mathematically find parameter values for which a generating partition is Markov and the RNG behaves as a Markov information source, and then to mathematically analyze the information generation process and the RNG. In the companion paper we discuss practical aspects of our chaos-based RNGs.

Index Terms—Chaos, random number generator, symbolic dynamics.

I. INTRODUCTION

RANDOM number generators (RNGs) are useful in every scientific area which uses Monte Carlo methods [1]. It is difficult to imagine a scientific area where Monte Carlo methods and RNGs are not used. Extremely important is the application of RNGs in cryptography for generation of cryptographic keys, and random initialization of certain variables in cryptographic protocols.

The choice of the RNG for a specific application depends on the requirements specific to the given application. If the ability to regenerate the random sequence is of crucial significance such as debugging simulations, or the randomness requirements are not very stringent (flying through space on your screen saver), or the hardware generation costs are unjustified, then one should resort to pseudo-random number generators (PRNGs). PRNGs are algorithms implemented on finite-state machines and are capable of generating sequences of numbers which appear random-like from many aspects. Though they are necessarily periodic (“Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin”, John von Neumann), their periods are very long, they pass many statistical tests and can be easily implemented with simple and fast software routines. However, when ultimate security is necessary one must turn to the only cipher which is theoretically unbreakable—one-time pad [2]. This cipher

requires a truly random sequence, and PRNGs are inappropriate for such a purpose. It is also an absolute necessity that cryptographic keys and initialization variables in cryptographic protocols are generated by RNGs. Otherwise, if a PRNG is employed, then security of the cryptographic algorithm and protocol can be no higher than the security of the PRNG. So, in all these cases where PRNGs are not suitable and unpredictability is a more important requirement than repeatability, one must turn to generators of truly random numbers. In the remaining part of the paper we will use the notion RNG to denote solely the generators of truly random numbers.

It is widely accepted that the core of any RNG must be an intrinsically random physical process. So, it is no surprise that the proposals and implementations of RNGs range from tossing a coin, throwing a dice [3], drawing from a urn, drawing from a deck of cards and spinning a roulette to measuring thermal noise from a resistor and shot noise from a Zener diode or a vacuum tube [4]–[9], measuring radioactive decay from a radioactive source [4]–[6], [10], integrating dark current from a metal insulator semiconductor capacitor [11], detecting locations of photoevents [12], and sampling a stable high-frequency oscillator with an unstable low-frequency clock [13]–[15]. There exist certain methods to convert the assumed randomness of a physical process into a sequence of discrete random variables (desirably independent and with identical distribution), most usually binary ones, and later on to derive the desired distribution from them. These methods suffer from the random and uncontrollable appearance of the random physical process, and consequently introduce biases in the binary sequence. In order to reduce any biases in the produced distribution, postprocessing of the produced sequence on a digital computer is usually done. Finally, the proper design and correct work (no silent breakdowns) of the RNG, and the assumed randomness of the physical process are checked via extensive statistical tests. However, one should keep in mind that no finite number of statistical tests can prove that a sequence is random, tests can only show that a sequence is not random.

Theory and tools of nonlinear systems and their chaotic behavior have provided an alternative and qualitatively different type of RNGs. Several authors have already proposed to use chaotic systems as sources of physical randomness [16]–[23]. When using chaotic systems, there is no need to assume the randomness, since when observed in a coarse-grained state-space they do behave randomly. However, the existing designs of chaotic RNGs still exhibit the same drawbacks as the classical RNGs based on the assumed randomness of a physical process.

To construct an RNG based on chaos, in this paper we exploit the double nature of chaos, deterministic in microscopic space

Manuscript received January 7, 2000; revised September 17, 2000. This work was supported in part by the Army Research Office under Grant DAAG55-98-1-0269, in part by the Department of Electronics under Grant DE-FG03-95ER14516, and in part, by the National Science Foundation under Grant NCR-9612250. This paper was recommended by Associate Editor C. K. Tse.

T. Stojanovski is with the Expert Information Services, Melbourne 3027, Australia (e-mail: tonis@expert.com.au).

L. Kocarev is with the Institute for Nonlinear Science, University of California, San Diego, La Jolla, CA 92093-0402 USA (e-mail: kocarev@heisenberg.ucsd.edu).

Publisher Item Identifier S 1057-7122(01)02202-4.

and by its defining equations, and random in macroscopic space. For our RNG, we can understand the mechanism of information source and can analytically analyze it in terms of Markov source state transition probabilities. Therefore, for our chaotic RNG and for any other chaotic RNG whose information generation mechanism is completely understood and analyzed, *in principal at theoretical level* there is no need for statistical tests. Our chaos-based RNG is mathematically proven to act as an information source, is not prone to silent breakdowns, its optimum parameters can be found, and it can be efficiently implemented on-chip.

Here is the layout of our paper. In Section II-A we define the notion of RNG. Section II-B introduces some notions from the theory of chaos which are necessary for the remaining discussion, and mathematically describes randomness and unpredictability of deterministic chaos. Section III shows the way a RNG based on a chaotic map can be analyzed. Concluding remarks are given in Section IV.

II. CHAOS AND RNGS

Theory of chaos, as a branch of theory of nonlinear dynamical systems, has brought to our attention a somewhat surprising fact, low-dimensional dynamical systems are capable of complex and unpredictable behavior. Complexity of defining equations of the source is not a necessary condition for generation of a random sequence. Making the RNG susceptible to influences from many independent random influences, as in [15], is not a necessity.

A. Definition of RNG

Definition 1: An ideal RNG is a discrete *memoryless* information source (DMIS) that generates *equiprobable* symbols. An RNG is a discrete information source with positive entropy. \square

An RNG is desired to generate a sequence of independent, identically distributed random variables. As shown later on several causes make it extremely difficult to implement an ideal RNG. A practical RNG behaves as an information source with memory and generates nonequiprobable symbols. Whenever appropriate, we will use the term *biased RNG* to distinguish it from the notion of ideal RNG as defined in Def. 1.

Quality of a biased RNG will be measured through its redundancy $\rho = \log_2 Q - h$ where Q and h are cardinality and entropy (to be defined later on in Section II) of the corresponding information source, respectively. Redundancy of an ideal RNG is equal to 0, and a nonzero redundancy of a biased RNG is a measure of its deviation from an ideal RNG. As an illustration of the usefulness of redundancy, we give an example. If an RNG with redundancy ρ generates cryptographic keys with length N bits, then an enemy will search on average $2^{(1-\rho)N}$ keys before finding the right key and an effective length of the keys can be defined as $N_e = (1 - \rho)N$.

B. Randomness in Deterministic Chaos

Throughout the paper, we will be considering only *deterministic chaotic* discrete time dynamical systems

$$\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n); \quad \mathbf{x} \in \mathcal{S} \subseteq \mathcal{R}^N \quad (1)$$

whose defining maps $\mathbf{f} : \mathcal{S} \rightarrow \mathcal{S}$ contain no random terms of any nature (noisy perturbations of the system's state, random variations of the parameter values etc.). We examine only dynamical systems whose evolution is determined by the defining vector field and the initial condition \mathbf{x}_1 . However, to completely specify an initial condition an infinite amount of information and a measuring system with an infinite precision are required, which are both intractable. What are the effects of a measuring system's finite precision?

Measuring an initial (and future) state is equivalent to partitioning the state space into a finite number of regions, and observing the evolution in this macroscopic world. Any set of m disjoint regions $\beta = \{\mathcal{C}_1, \dots, \mathcal{C}_m\}$ which covers the state space \mathcal{S} of (1) is called a partition of (1), that is

$$\beta = \{\mathcal{C}_1, \dots, \mathcal{C}_m\} \quad \bigcup_{i=1}^{i=m} \mathcal{C}_i = \mathcal{S} \quad \mathcal{C}_i \cap \mathcal{C}_j = \emptyset, \quad \text{for } i \neq j.$$

The partition with $m = 1$ is called trivial partition. For a partition β , we denote the union of boundaries between regions of β as $\mathcal{B}(\beta)$. If we allow for the regions of β to overlap, then the set β is called open cover of \mathcal{S} .

A unique symbol $i = \sigma(\mathcal{C}_i), i \in M = \{1, 2, \dots, m\}$ is assigned to every region $\mathcal{C}_i \in \beta$. The process of partitioning the state space, assigning symbols to every region from the partition, and the resulting macroscopic dynamics are called symbolic dynamics. Denote with $\Psi = \prod_{j=1}^{\infty} M$, the space of all sequences $X_1^\infty = X_1 X_2 \dots X_j \dots$ with infinite length, where $X_j \in M$. This way, we obtain a map $\mu_\beta : \mathcal{S} \rightarrow \Psi$ defined as

$$\mu_\beta(\mathbf{x}_1) = X_1^\infty \Leftrightarrow \mathbf{f}^{j-1}(\mathbf{x}_1) \in \mathcal{C}_{X_j} \Leftrightarrow \mathbf{x}_1 \in \bigcap_j \mathbf{f}^{-j+1}(\mathcal{C}_{X_j}), \quad \text{for } j \geq 1$$

which assigns a sequence $X_1^\infty \in \Psi$ to every point $\mathbf{x}_1 \in \mathcal{S}$, and X_j is the symbol generated at time j . Since (1) is chaotic, $\mathbf{f}^j(\mathcal{C}_i), j > 1$, may expand over several regions for some $\mathcal{C}_i \in \beta$. Different initial states belonging to the same region \mathcal{C}_{X_1} will produce different observations at some later time $j > 1$. From the viewpoint of our measuring system, identical macroscopic initial states evolve differently. A loss of determinism occurred, and transitions between the regions of β can only be specified by means of probabilities. Partitioning of the state space turns the deterministic chaotic system (1) into an ergodic information source which can be analyzed in terms of information theory. Ergodicity of the source follows from the assumption that (1) has a single chaotic attractor, that is, from the ergodicity of its invariant measure [25]. The source tends to become stationary for mixing maps, in which case every initial measure leads to the ergodic invariant measure. For the newly obtained information source one can compute entropies

$$H_n^\beta = - \sum_{X_1^n} P(X_1^n) \log P(X_1^n)$$

with $P(X_1^n)$ being the probability of occurrence of trajectory subsequence (word) X_1^n . H_n^β quantifies the average uncertainty when predicting words of length n . Throughout the paper, we

use logarithms with base 2 and the amount of information will be expressed in bits. The conditional entropy of the $(n + 1)$ -th symbol in the macroscopic trajectory when the previous n symbols are known is equal to

$$h_n^\beta = H_{n+1|n}^\beta = \begin{cases} H_{n+1}^\beta - H_n^\beta, & \text{for } n \geq 1 \\ H_1^\beta, & \text{for } n = 0 \end{cases}$$

Source entropy of (1) for a partition β is defined by

$$h^\beta = \lim_{n \rightarrow \infty} h_n^\beta = \lim_{n \rightarrow \infty} \frac{1}{n} H_n^\beta.$$

Kolmogorov-Sinai (KS) entropy of (1) is the supremum of the source entropy over all possible partitions

$$h_{\text{KS}} = \sup_{\beta} h^\beta.$$

If $h^\beta = h_{\text{KS}}$, then β is a generating partition. An interesting property of a generating partition β is that the corresponding map μ_β is injective, that is, $\mathbf{x}' \neq \mathbf{x}'' \Rightarrow \mu_\beta(\mathbf{x}') \neq \mu_\beta(\mathbf{x}'')$.

Turning a deterministic chaotic system into an information source via partitioning of the state-space is not in collision with Shannon's note [26] that a deterministic system can not generate information. Actually, a chaotic system does not generate information, that is, its evolution is completely determined by its initial state $H(\mathbf{x}_n | \mathbf{x}_1) = 0$. A chaotic system merely converts the information about its initial state into a form which is visible to the measuring system. Every letter in the coarse-grained trajectory, which is a sequence of letters, brings additional amount of information about the initial state. To explain this, we define a refinement β^n at stage n for a given partition β as consisting of the following regions:

$$\mathcal{C}_{X_1^n} = \left\{ \mathbf{x} \in \mathcal{S} \mid \mathbf{x} \in \bigcap_{j=1}^n \mathbf{f}^{-j+1}(\mathcal{C}_{X_j}) \right\}, \quad X_1^n \in M^n$$

where $\mathbf{f}^{-j+1}(\mathcal{C}_{X_j}) = \{\mathbf{x} \in \mathcal{S} \mid \mathbf{f}^{j-1}(\mathbf{x}) \in \mathcal{C}_{X_j}\}$, for $j = 1, \dots, n$. For a particular initial state \mathbf{x}_1 the first n symbols X_1^n specify one region of β^n to which \mathbf{x}_1 belongs. The next symbol X_{n+1} brings a positive amount of information h_n about \mathbf{x}_1 , and points to the region of β^{n+1} which contains \mathbf{x}_1 . This additional information is expressed through the fact that $\text{diam}(\beta^{n+1}) < \text{diam}(\beta^n)$ for $n \geq 1$ with $\text{diam}(\beta^n)$ being the maximum diameter of all the regions that β^n consists of. This means that every new symbol X_{n+1} from the sequence $\mu_\beta(\mathbf{x}_1)$ specifies \mathbf{x}_1 with higher and higher precision. If β is a generating partition, then $\lim_{n \rightarrow \infty} \text{diam}(\beta^n) = 0$.

Now, the way a chaotic system can be used as an RNG is obvious: find a partition β of the state-space that produces a DMIS. There are very simple chaotic maps that satisfy the requirements for an RNG at theoretical level. Amongst the simplest chaotic maps which satisfy the requirements are, the logistic map $x_{n+1} = rx_n(1 - x_n)$ with fully developed chaos $r = 4$, and its homeomorphically conjugated maps such as Bernoulli shift, tent map etc. Following binary generating partition $\beta = \{[0, 1/2), [1/2, 1)\}$, produces a binary DMIS.

III. ANALYSIS OF CHAOS-BASED RNGS

A. Simplicity of Piecewise Linearity

Current knowledge in chaos theory allows us to rigorously analyze only very simple dynamical systems, while for high-dimensional systems or those with complicated nonlinearity one must rely on numerical analysis. The advantages of chaos-based RNGs over classical ones emerge from the existence of mathematical tools for analysis of chaotic systems, and particularly from their information generation mechanism. Therefore, when one designs a chaos-based RNG, one should stick to the simplest to analyze and understand chaotic systems. As shown later on, this ability is the key for the optimum design of an RNG. Consequently, it is no surprise that we choose and analyze a class of PL1D maps defined with

$$x'_{n+1} = \begin{cases} q'_1 + k'_1(x'_n - T_L), & \text{for } x'_n < T_L \\ q'_2 + k'_2(x'_n - T_L), & \text{for } x'_n \geq T_L \end{cases} \quad (2)$$

where $k'_1, k'_2 > 1, q'_1 > T_L > q'_2$. As map (2) is expanding everywhere, there are no micro Feigenbaum diagrams (there are no stable periodic windows) in the chaotic region.

Why do we consider PL1D maps as being very simple? 1) A 2-regions PL1D map can be with $h_{\text{KS}} = 1$ [bit], which is sufficient for a binary RNG. 2) $T_L \in \mathcal{B}(\beta) \Leftrightarrow \beta$ is a generating partition. 3) The number of parameters is very small and analysis of sensitivity of map's properties on parameters' variations can be analytically attainable. 4) PL1D maps can be simply implemented by virtue of switched capacitor [16], [18] and switched current circuits [20], which can operate at high frequencies.

PL1D maps with more linear regions can have higher h_{KS} , which is desirable, but this is at the expense of additional hardware: more threshold levels, more logical circuits to determine which region the current state belongs to, and of increased number of parameters.

How about other types of nonlinearity? Any other type of nonlinearity is much more difficult to analyze. For example, the logistic map contains only a quadratic term and we are still unable to analytically find its natural invariant measure, metric and topological entropy except for the case of fully developed chaos $r = 4$. Finding the shape of redundancy in the macroscopic evolution for $r < 4$ is beyond the reach of current knowledge, while this is a tractable task for PL1D maps. Finding generating partitions for higher-dimensional maps is a difficult task. Practical implementation of logistic map and other maps with more complicated nonlinear terms is a difficult task too. The multiplier of analog signals from the practical realization of the logistic map in [27] operates at frequencies lower than switching circuits from [16], [18], [20].

So, two-region PL1D maps serve not only as a paradigm for chaotic RNGs, but also as the most appropriate maps for RNG purposes.

B. Linear Conjugacy

For every set of parameters of map (2), following transformation

$$x = \begin{cases} (x' - T_L)/(T_L - q'_2), & \text{for } k'_1 \leq k'_2 \\ (x' - T_L)/(T_L - q'_1), & \text{for } k'_1 > k'_2 \end{cases} \quad (3)$$

yields a linearly conjugate map

$$x_{n+1} = f(x_n) = \begin{cases} q_1 + k_1 x_n, & \text{for } x_n < 0 \\ -1 + k_2 x_n, & \text{for } x_n \geq 0 \end{cases} \quad (4)$$

where parameters of (2) and (4) are related via

$$\begin{cases} k_1 = k'_1 & k_2 = k'_2, & q_1 = \frac{q'_1 - T_L}{T_L - q'_2}, & \text{for } k'_1 \leq k'_2 \\ k_1 = k'_2 & k_2 = k'_1, & q_1 = \frac{T_L - q'_2}{q'_1 - T_L}, & \text{for } k'_1 > k'_2 \end{cases} \quad (5)$$

Due to the linear conjugacy between (2) and (4), map (4) retains the entropies, Lyapunov exponent, Markov character of partitions (to be described later on), and almost all other features of (2). Reduction in the number of parameters from five to three, results in a simpler analysis and better understanding of (4) than that of map (2). The fact is that $k_1 \leq k_2$ further decreases the required number of sets of parameters for a given precision. From now on, our discussion will relate to map (4).

C. Parasitic Attractors

The most notable representative of the family of maps (4) is

$$x_{n+1} = \begin{cases} 1 + 2x_n, & \text{for } x_n < 0 \\ -1 + 2x_n, & \text{for } x_n \geq 0 \end{cases} \quad (6)$$

where $k_1 = k_2 = 2, q_1 = 1$. If it were possible to implement (6), then the problem of RNG would be immediately and easily solved, as mentioned at the end of Section II-B. However, the following problems are present. Map (6) is chaotic on the interval $(-1, 1)$ only. Initial states out of this interval move quickly to either $-\infty$ or $+\infty$. Map (6) is ergodic on $(-1, 1)$ and can come arbitrarily close to -1 and 1 , when only a small noise is sufficient to shift the map's state out of $(-1, 1)$. Therefore, the ergodic invariant set $(-1, 1)$ is not an attractor. Moreover, the assumption of fluctuating parameters which may cause slopes larger than 2 is plausible in practical implementation of (4). For slopes $k_1 = k_2 > 2$, the measure of initial states from $(-1, 1)$ which remain in $(-1, 1)$ after n iterations exponentially decreases to 0 with n , and map (4) exhibits chaotic motion on a Cantor set with zero measure. So, almost all initial states from $(-1, 1)$ are attracted to either $-\infty$ or $+\infty$. In order to make the chaotic motion attracting, one may employ Bernoulli shift $x_{n+1} = 2x_n \bmod 1$ in which case $(0, 1)$ is a global attractor. Actually, only two additional linear regions as in

$$x_{n+1} = \begin{cases} 3 + 2x_n, & \text{for } x_n < -1 \\ 1 + 2x_n, & \text{for } -1 \leq x_n < 0 \\ -1 + 2x_n, & \text{for } 0 \leq x_n < 1 \\ -3 + 2x_n, & \text{for } x_n \geq 1 \end{cases} \quad (7)$$

would serve the purpose, but this requires additional threshold circuits and more complex hardware, and entangles the analysis.

In practical implementations of map (4), the maximum and minimum values of the map's states are limited by saturation. This introduces regions of constant output values in map (4) as illustrated on Fig. 1. The chaotic attractor is bounded to $(-1, q_1)$. If the map does not intersect with the line $x_{n+1} = x_n$, then there are no attracting points. When an attracting point exists, for example, point P from Fig. 2, then the basin of attraction of the chaotic attractor is $(-\infty, U_2)$ and does not include value I_+ corresponding to the positive saturation value. U_2 is

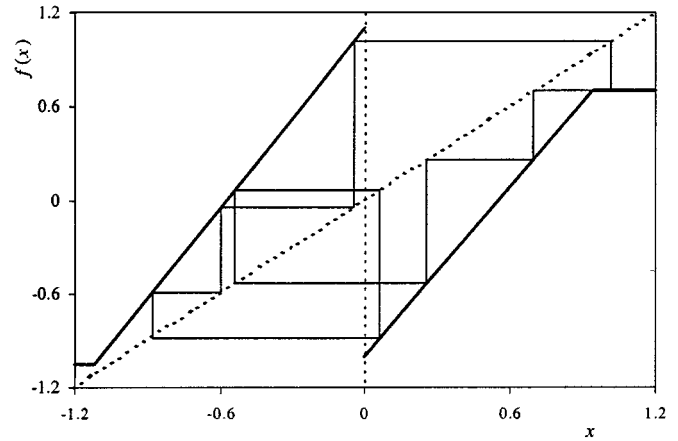


Fig. 1. Map (4) for $k_1 < k_2 < 2$. Positive saturation value I_+ produces an attracting point P, while negative saturation value I_- does not.

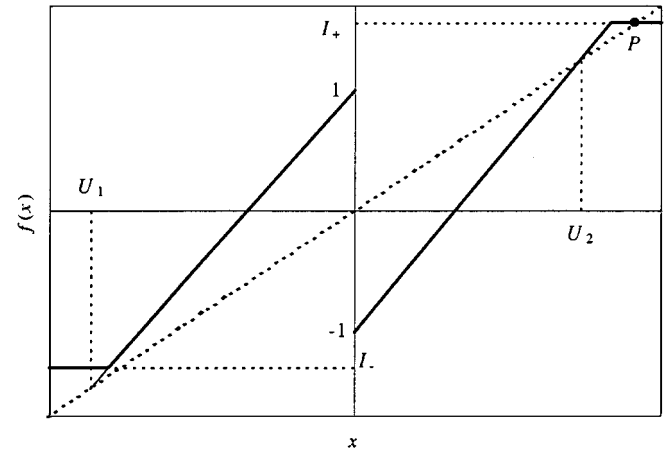


Fig. 2. Parasitic periodic orbit of period 8 (solid thin line) caused by a small positive saturation value $I_+ < f(q_1)$. Parameters of map (4) are $q_1 = 1.105, k_1 = 1.93, k_2 = 1.8, I_- = -1.054$ and $I_+ = 0.698$.

the intersection point of the lines $y = -1 + k_2 x$ and $y = x$, and is equal to $U_2 = 1/(k_2 - 1)$. Attracting point P is with basin of attraction $(U_2, +\infty)$. As a result, power-on transient may lead to a parasitic stable point instead to the desired chaotic motion. Even when the power-on transient leads to the chaotic attractor, if $f(q_1)$ is very close to U_2 , then a noise larger than $U_2 - f(q_1)$ will force the map to leave the chaotic attractor and settle on the point attractor P. A parasitic point attractor will appear unless $I_+ < U_2$. On the other hand, it is a mandatory requirement that $I_+ > f(q_1)$. Otherwise, a periodic attractor appears instead of the intended chaotic attractor as illustrated on Fig. 2. In the sequel, we use the term parasitic periodic attractors to denote both point and periodic attractors. In order to assure a reliable operation of a chaos-based RNG, it is a must that the chaotic attractor is with the global basin of attraction.

Hence, the region of allowed values for I_+ is $(f(q_1), U_2)$ and has a length, $k_2/(k_2 - 1) - k_2 q_1$. The maximum allowed change in I_+ which does not cause a parasitic attractor is given by $L_+ = (U_2 - f(q_1))/(U_2 + f(q_1)) = (k_2/(k_2 - 1) - k_2 q_1)/(k_2 q_1 - 1 + 1/(k_2 - 1))$, where the numerator and denominator are length and median, respectively, of the region of allowed values for I_+ . The derivative of L_+ with k_2 is negative for $k_2 > (1 + q_1)/(2q_1)$, and L_+ decreases for larger k_2 . If k_2 is smaller than

but very close to 2 and q_1 is close to 1, then L_+ is very small. Then, even if the nominal value of I_+ assures a global chaotic attractor, a small increase or decrease in I_+ due to a temperature or a power supply fluctuation may destroy the chaotic motion. Similar problems arise when $k_1 \approx 2$. The region of allowed values for I_- is $(U_1, f(-1))$, where $U_1 = -q_1/(k_1 - 1)$, and is with length $q_1 k_1 / (k_1 - 1) - k_1$. The maximum allowed change in I_- is $L_- = |U_1 - f(-1)| / |U_1 + f(-1)| = (q_1 k_1 / (k_1 - 1) - k_1) / (k_1 - q_1 + q_1 / (k_1 - 1))$. Again, L_- decreases for larger k_1 when $k_1 > (q_1 + 1)/2$. I_+ and I_- are set to the median of their allowed regions. For example, $L_+ = 0.2$ means that I_+ can change $\pm 20\%$ from its nominal value without producing a parasitic attractor. The margin against appearance of parasitic attractors is $L = \min(L_+, L_-)$, where we implicitly assume that variations in I_+ and I_- are with same extent.

Parameters I_- and I_+ influence the steady behavior of (4) only when parasitic attractors exist. Once I_- and I_+ are such that there are no parasitic attractors, I_- and I_+ can be omitted from the analysis of (4). In the sequel, in order to keep (4) simple, we will not include description of the constant regions in it, though we will keep in mind that these regions always exist.

From (3) and the need to avoid parasitic attractors ($L_+, L_- > 0$), it follows that behavior of (4) should be analyzed only in the region $\mathcal{P} = \{(k_1, k_2, q_1) \mid 1 < k_1 < 2, k_1 \leq k_2 < 2, k_1 - 1 < q_1 < 1/(k_2 - 1)\}$ of the 3-D parameter space $k_1 \times k_2 \times q_1$.

D. Generating and Markov Partitions

In the sequel, we will consider a binary generating partition $\beta = \{\mathcal{C}_1, \mathcal{C}_2\}$ only, where $\mathcal{C}_1 = (q_2, 0)$ and $\mathcal{C}_2 = [0, q_1]$. Thus, we implicitly assume that there is no mismatch between the boundary point 0 of β and the discontinuity point 0 of the PL1D map. This assumption is justified by the practical implementation of the PL1D map and the RNG, where a single threshold circuit is used to both iterate the map, that is, to implement the discontinuity point 0 and to generate output bits, that is, to implement boundary point 0 of β . Using a single threshold circuit also implies simpler hardware.

Clearly, we are interested in the properties of the information source emerging from the tapping of (4) in the β partitioned state space. Is it a Markov chain? If yes, then what is the order of the Markov chain? And what is the dependence of the order of the Markov chain on the parameters? Are the parameters which give rise to a Markov chain dense in the parameter region \mathcal{P} ? Answers to these questions will also provide a clue to the source and shape of the redundancy contained in the binary sequence. This way, we can evaluate cost of the task of removing this redundancy in terms of required hardware and reduction of the bit generation rate. Such an analysis will also reveal computational complexity of the task to an intruder who wants to recognize this redundancy and benefit from it e.g., by reducing the searching of the key space, is faced with.

Next, we define the notion of a Markov partition.

Definition 2: A partition β is a Markov partition (of order 1) if the boundaries $\mathcal{B}(\beta)$ are kept invariant by the dynamics

$$f(\mathcal{B}(\beta)) \subseteq \mathcal{B}(\beta), \square \quad (8)$$

In other words, every region $\mathcal{C}_i \in \beta$ is mapped to a union of regions from β , that is, $f(\mathcal{C}_i) \cap \mathcal{C}_j \neq \emptyset \Rightarrow f(\mathcal{C}_i) \supseteq \mathcal{C}_j$. The notion of Markov partition can be refined to higher orders.

Definition 3: A partition β is a Markov partition of order $r > 1$ if

$$f(\mathcal{B}(\beta^{r-1})) \setminus \mathcal{B}(\beta^{r-1}) \neq \emptyset \quad \text{and} \quad f(\mathcal{B}(\beta^r)) \subseteq \mathcal{B}(\beta^r) \quad (9)$$

that is, β^{r-1} is not a Markov partition and β^r is a Markov partition. \square

In the remaining part of the paper, the term Markov partition will be used to denote a Markov partition of any finite order. If the successive states of the coarse grained dynamics are independent, then the partition is called Bernoulli partition. Markov character of a partition is only a topological notion, and though it is a necessary condition, it does not imply Markov character of the information source in general. When the information source is Markov, then its order is less than or equal to the order of the Markov partition. If a Markov partition exists for a given map, then the map is called Markov.

The main motivation to search for Markov partitions is presented next. The following discussion will also support our decision to insist on piecewise linear maps. There is no general way to analytically find the natural invariant density using Perron-Frobenius operator, and then to compute KS entropy or entropy for a given partition. This problem is highly relieved and analytically tractable when the chaotic information source is Markov. Piecewise linear maps which are linear inside each of the regions of the Markov partition give rise to a Markov source [28], [29]. Their natural invariant density is piecewise constant, and Perron-Frobenius operator can be substituted by the transition stochastic matrix of the Markov source whose transition probabilities are [30]

$$P_{ij} = \frac{\mathcal{L}(\mathcal{C}_j \cap f^{-1}(\mathcal{C}_i))}{\mathcal{L}(\mathcal{C}_j)} \quad (10)$$

where $\mathcal{L}(\cdot)$ denotes Lebesgue measure. One can analytically find transition probabilities P_{ij} via (10), state probabilities via inverting or iterating the transition matrix, natural invariant density via dividing region's probabilities by region's Lebesgue measures, and structure and amount of information redundancy. We are not aware of work showing Markov character of symbolic dynamics for other families of Markov maps other than piecewise linear ones.

From the piecewise linearity of (4) an additional crucial consequence follows. Namely, if x and its next $n - 1$ iterations belong to a same linear region, then x and $f^n(x)$ are simply related. If $f^i(x) < 0$ for $i = 0, 1, \dots, n - 1$, then $f^n(x) = q_1 \sum_{i=0}^{n-1} k_1^i + k_1^n x$. If $f^i(x) \geq 0$ for $i = 0, 1, \dots, n - 1$, then $f^n(x) = -1 \sum_{i=0}^{n-1} k_2^i + k_2^n x$. As shown later on, this simplicity of the composition of (4) enables us to easily search for Markov maps.

E. Dependence on Parameters

Smaller values for k_1 and k_2 give a larger margin against complete failure in sense of abandoning the chaotic motion. On the other hand, values for k_1 and k_2 closer to 2 give higher entropy

where $S_0 = 0$, $S_i = S_{i-1} + J_i$ for $i = 1, \dots, m$, and $r = S_m - 1$. Substituting $\sum_{j=0}^{J_i-1} k^j = (k^{J_i} - 1)/(k - 1)$ we get

$$\begin{aligned} & (-1)^m k^{S_m} (k - 1) + \sum_{i=0}^{m-1} (-1)^i k^{S_i} (k^{J_{i+1}} - 1) \\ &= (-1)^m k^{S_m} (k - 1) + \sum_{i=0}^{m-1} (-1)^i (k^{S_{i+1}} - k^{S_i}) = 0 \end{aligned}$$

After some transformations, we get

$$(-1)^m k^{S_m+1} - 2 \sum_{i=1}^m (-1)^i k^{S_i} - 1 = 0$$

and the proof is finished. \square

The set of k values which produce Markov partitions is a countably infinite set, and therefore is with Lebesgue measure 0. Though improbable in practice, these k values are dense in \mathcal{P}_1 , and information generation mechanism can be analytically analyzed arbitrarily close to any point from \mathcal{P}_1 .

For $1 < k \leq \sqrt{2}$ the map is still chaotic but its attractor is not the whole interval $(-1, 1)$ any more. One can easily show that for $1 < k \leq \sqrt{2}$, where $f_1^2(-1) > f_1(1)$, chaotic attractor is bounded to $[-1, f_1^2(1)] \cup [f_1(-1), f_1(1)] \cup [f_1^2(-1), 1]$. While the intervals $[L_-, -1)$ and $[1, L_+]$ are attracted to the attractor in a finite (small) number of iterations, the measure of points from $A = [f_1^2(1), f_1(-1)] \cup [f_1(1), f_1^2(-1)]$ which after n iterations are still in A exponentially decreases with n , until eventually only the two fixed points of f_1^2 , $\pm 1/(k+1)$, remain in A . The map $f_1^2(x)$ can be decomposed into two maps $f_{1a}^2(x)$ and $f_{1b}^2(x)$ which are both conjugate to $f_1(x)$ with slope k^2 . This proves a nonmixing behavior of the map for $1 < k \leq \sqrt{2}$. If β is a Markov partition of order r for slope k , then for slope \sqrt{k} the order is $2r + 1$. For $1 < k \leq \sqrt{2}$ symbolic sequences for points on the chaotic attractor consist of binary pairs 01 and 10. Such a structured redundancy can be easily removed by skipping every second bit, and this way one obtains a binary sequence with entropy $\log k^2$. Skipping every second bit is equivalent to obtaining the bits from either $f_{1a}^2(x)$ or $f_{1b}^2(x)$. β is generating partition for $f_1(x)$, $f_{1a}^2(x)$ and $f_{1b}^2(x)$. For $\sqrt[3]{2} < k \leq \sqrt{2}$, Markov chain is a periodic persistent group of order 2. This discussion can be easily generalized for $\sqrt[2n+1]{2} < k \leq \sqrt[2n]{2}$ e.g., $f_1^{2n}(x)$ is decomposable into 2^n maps which are all linearly conjugate to $f_1(x)$ with slope k^{2^n} .

On basis of the results presented in this sub-section we conjecture that sensitive dependence of Markov partition's order on parameter changes is typical for chaotic maps. In other words, Markov character of a partition is not a generic property and a small perturbation of the chaotic map will destroy the Markov character. This sub-section leads to another conclusion, the set of parameters for which β is a Markov partition is dense in the parameter region \mathcal{P} . We give following plausible explanation. Condition (8) reduces to the requirement that q_1 and -1 belong to the basins of attraction of two periodic orbits. From the ergodicity of map (4) it follows that this condition can be satisfied for any region in \mathcal{P} .

IV. CONCLUSION

An RNG based on a 2-regions PLID chaotic map is analyzed, which overcomes the drawbacks of classical RNGs: reliance on the assumed randomness of a physical process, inability to analyze and optimize the RNG, inability to compute probabilities and entropy of the RNG, and inconclusiveness of statistical tests. We exploit the deterministic part of the nature of chaos and the simplicity of the 2-regions PLID map to extend existing works on chaos-based RNGs, and to give mathematical analysis of the information generation mechanism.

REFERENCES

- [1] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia, PA: SIAM, 1992.
- [2] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, no. 3, pp. 656–715, 1949.
- [3] S. M. Mathyas and C. H. Meyer, "Generation, distribution and installation of cryptographic keys," *IBM Syst. J.*, vol. 17, no. 2, pp. 126–137, 1978.
- [4] C. H. Vincent, "The generation of truly random binary numbers," *J. Physics E*, vol. 3, no. 6, pp. 594–598, 1970.
- [5] —, "Precautions for accuracy in the generation of truly random binary numbers," *J. Physics E*, vol. 4, no. 9, pp. 825–828, 1971.
- [6] R. S. Maddocks, S. Matthews, E. W. Walker, and C. H. Vincent, "A compact and accurate generator for truly random binary digits," *J. Physics E*, vol. 5, no. 5, pp. 542–544, 1972.
- [7] H. F. Murry, "A general approach for generating natural random variables," *IEEE Trans. Comput.*, vol. 19, pp. 1210–1213, 1970.
- [8] N. O. Sokal, "Optimum choice of noise frequency band and sampling rate for generating random binary digits from clipped white noise," *IEEE Trans. Comput.*, vol. 21, pp. 614–615, 6 1972.
- [9] W. T. Holman, J. A. Connelly, and A. B. Dowlatabadi, "An integrated analog/digital random noise source," *IEEE Trans. Circuits Syst. I*, vol. 44, pp. 521–528, June 1997.
- [10] J. Walker, *HotBits: Genuine random numbers, generated by radioactive decay* [Online]. Available: <http://www.fourmilab.ch/hotbits/>
- [11] G. B. Agnew, "Random sources from cryptographic systems," in *Advances in Cryptology—CRYPTO'85*. New York: Springer-Verlag, 1986, pp. 77–81.
- [12] A. J. Martino and G. M. Morris, "Optical random number generator based on photoevent locations," *Appl. Opt.*, vol. 30, no. 8, pp. 981–989, 1991.
- [13] Rand Corporation, *A Million Random Digits with 100 000 Normal Deviates*. Glencoe, IL: Free Press, 1955.
- [14] R. C. Fairfield, R. L. Mortenson, and K. B. Coulthart, "An LSI random number generator (RNG)," in *Advances in Cryptology—Crypto'84*. New York: Springer-Verlag, 1984, pp. 203–230.
- [15] L. Letham, D. Hoff, and A. Folmsbee, "A 128 K EPROM using encryption of pseudorandom numbers to enable random access," *IEEE J. Solid-State Circuits*, vol. SC-21, pp. 881–887, Oct. 1986.
- [16] S. Espejo-Meana, A. Rodriguez-Vazquez, J. L. Huertas, and J. M. Quintana, "Application of chaotic switched-capacitor circuits for random-number generation," in *Proc. Eur. Conf. Circuit Theory and Design 1989*, 1989, pp. 440–444.
- [17] G. M. Bernstein and M. A. Lieberman, "Secure random number generation using chaotic circuits," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 1157–1164, Dec., 1990.
- [18] A. Rodriguez-Vazquez, M. Delgado, S. Espejo, and J. L. Huertas, "Switched-capacitor broadband noise generator for CMOS VLSI," *Electron. Lett.*, vol. 27, no. 21, pp. 1913–1915, 1991.
- [19] M. Delgado-Restituto, A. Rodriguez-Vazquez, S. Espejo, and J. L. Huertas, "A chaotic switched-capacitor circuit for 1/f noise generation," *IEEE Trans. Circuits Syst. I*, vol. 9, pp. 325–328, Apr. 1992.
- [20] M. Delgado-Restituto, F. Medeiro, and A. Rodriguez-Vazquez, "Non-linear switched-current CMOS IC for random signal generation," *Electron. Lett.*, vol. 29, no. 25, pp. 2190–2191, 1993.
- [21] T. Kuusela, "Random number generation using a chaotic circuit," *J. Non-linear Sci.*, vol. 3, no. 4, pp. 445–458, 1993.
- [22] D. Davis, R. Ihaka, and P. Fenstermacher, "Cryptographic randomness from air turbulence in disk drives," in *Advances in Cryptology—CRYPTO'94*. New York: Springer-Verlag, 1994, pp. 114–120.

- [23] B. Vizvári and G. Kolumbán, "Quality evaluation of random numbers generated by chaotic circuits for secure communication," Rutcor Research Rep. RRR 13–96, Apr. 1996.
- [24] H. G. Schuster, *Deterministic Chaos: An Introduction*. New York: VCH, 1989.
- [25] J. P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," *Rev. Mod. Phys.*, vol. 57, no. 4, pp. 617–656, 1985.
- [26] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
- [27] G. C. McGonigal and M. I. Elmasry, "Generation of noise by electronic iteration of the logistic map," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 981–983, Sept. 1987.
- [28] G. Györgyi and P. Szépfalussy, "Calculation of the entropy in chaotic systems," *Phys. Rev. A*, vol. 31, no. 5, pp. 3477–3479, 1985.
- [29] G. Nicolis and C. Nicolis, "Master equation approach to deterministic chaos," *Phys. Rev. A*, vol. 38, no. 1, pp. 427–433, 1988.
- [30] S. Grossmann and S. Thomae, "Invariant distributions and stationary correlation functions of one-dimensional discrete processes," *Z. Nat.*, vol. 32a, no. 10, pp. 1353–1363, 1977.
- [31] C. S. Hsu and M. C. Kim, "Method of constructing generating partitions for entropy evaluation," *Phys. Rev. A*, vol. 30, no. 6, pp. 3351–3354, 1984.
- [32] ———, "Construction of maps with generating partitions for entropy evaluation," *Phys. Rev. A*, vol. 31, no. 5, pp. 3253–3265, 1985.

Toni Stojanovski was born in 1968 in Skopje, Macedonia. He received the B.Sc. and M.Sc. degrees in electrical engineering, and communications, in 1990 and 1995, respectively, both from the Ss Cyril and Methodius University, Skopje, Macedonia, and the Ph.D. degree in communications from the RMIT University, Melbourne, Australia, in 1999.

Since 1998, he has been with the Expert Information Services, Melbourne, Australia. He has published more than 20 articles in journal and conference proceedings in the areas of chaos theory and communications.

Ljupčo Kocarev (SM'95) received the Ph. D degree in physics from the University Kiril i Metodij, Skopje, Macedonia, in 1989.

He is an Associate Research Scientist at the Institute for Nonlinear Science at the University of California, San Diego. His scientific interests include nonlinear science and its application to physics, biology and electrical engineering. He has authored more than 60 journal articles in the areas of chaos theory and communications, various international journals, including IEEE TRANSACTION ON CIRCUITS AND SYSTEMS I, and IEEE TRANSACTION ON CIRCUITS AND SYSTEMS II.