A Design for a Physical RNG with Robust Entropy Estimators

Wolfgang Killmann[1], Werner Schindler[2]

[1] T-Systems ISS GmbH
Rabinstr. 8
53111 Bonn, Germany
Wolfgang.Killmann@t-systems.com
[2] Bundesamt für Sicherheit in der Informationstechnik (BSI)
Godesberger Allee 185–189
53175 Bonn, Germany
Werner.Schindler@bsi.bund.de

Abstract. The development of a secure physical RNG is a two-fold challenge. First of all, an appropriate design has to be implemented carefully. Secondly, evidence shall be given that the design itself and in particular the concrete implementation are indeed secure, and this task may be even more challenging. We briefly address general aspects that reliable security evaluations of physical RNGs should take into account. We discuss a performant RNG design that is based on pairs of noisy diodes. We formulate and analyze a stochastic model that interestingly also fits to other RNG designs. We prove a theorem that provides tight lower bounds for the entropy per random bit, and we apply these results to a prototype of a physical RNG. Principally, this RNG could generate up to 700 000 random bits per second with average entropy $> 1 - 10^{-5}$.

Keywords: Physical RNG, stochastic model, entropy.

1 Introduction

Many cryptographic mechanisms require random numbers, e.g. as session keys, signature parameters, ephemeral keys (DSA, ECDSA), zero-knowledge protocols, challenge response-protocols, nonces. Inappropriate RNGs may totally weaken principally strong cryptosystems, e.g. if an adversary is able to determine session keys. Ideally, random numbers were uniformly distributed on their range and independent. However, this condition characterizes an *ideal RNG*, which is a mathematical construction (lastly a fiction). Following [10] (cf. [24] for further explanations) 'real-world' RNGs can be divided into two classes, which comprise the *true RNGs* (TRNGs) and the *deterministic RNGs* (DRNGs; aka pseudo-random number generators), respectively. The TRNGs fall into two subclasses: *Physical TRNGs* use non-deterministic effects of electronic circuits (e.g. shot noise from Zener diode, inherent semiconductor thermal noise, free running oscillators) or physical experiments (e.g., time between emissions of radioactive

decay, quantum photon effects). *Non-physical non-deterministic RNGs* exploit non-deterministic events (e.g., system time, hard disk seek time RAM content, user interaction). So-called *hybrid RNGs* combine design elements from both, TRNGs and DRNGs.

Unlike for deterministic RNGs it is hardly possible to specify approved designs for physical RNGs since security-relevant properties depend on the concrete implementation. A designer of a physical RNG is faced with two challenges. At first he has to develop an appropriate design and to implement it carefully. The second task may be even more difficult, namely to give evidence that the generic design and its implementation are indeed secure.

In the last years various designs of physical RNGs have been proposed [5–8, 14], and several evaluation guidances and standards were developed and became effective [1, 19, 2, 12, 10]. These documents define properties that strong RNGs should fulfil, and the evaluation guidances (cf. [1, 19, 2, 12]) explain how these criteria shall be validated. A comprehensive treatment of evaluation aspects for physical RNGs are given in [25].

In Section 2 we briefly address central aspects and goals that reliable security evaluations of physical RNGs should consider. In Section 3 we discuss an RNG design that exploits a pair of noisy diodes. In Section 4 we formulate and analyze a stochastic model that fits to this design and, interestingly, also to other RNG designs. In particular, we prove a theorem that allows to quantify a tight lower bound for the average entropy per random bit. We use our results to verify that a particular physical RNG might principally output up to 700 kBit random numbers with average entropy per random bit $> 1 - 10^{-5}$. We discuss a generic online test scheme that is tailored to RNG designs which to the analyzed stochastic model. The paper ends with final remarks.

2 Evaluation of Physical RNGs: Fundamental Aspects

In this section we address central aspects that are relevant for security evaluations of physical RNGs. For a comprehensive treatment of this topic we refer the interested reader to [24, 25, 21].

2.1 Entropy

With regard to Section 4 we extend the common definition of Shannon entropy from finite to (infinite) countable sets Ω (e.g. $\Omega = \mathbb{N}_0$). More precisely, to a random variable X that assumes values in a Ω we assign

$$H(X) := -\sum_{\omega \in \Omega} \mathrm{Prob}(X = \omega) \log_2(\mathrm{Prob}(X = \omega)). \tag{1}$$

According to the common convention we denote the Shannon entropy briefly as 'entropy' in the following.

Remark 1. (i) We point out that $H(X) \in [0, \infty]$ where $H(X) = \infty$ is principally possible for non-finite range of X. The 'auxiliary' random variables $V_{(s')}$ which will be of particular importance in Section 4) yet have finite entropy for any $s' \in (0, \infty)$ (cf. [22], Lemma 2(ii)).
(ii) Physical RNGs can usually be modelled by stationary stochastic processes (cf. Sect. 4). The internal random numbers (cf. Definition 1) typically assume values in $\Omega = \{0, 1\}$, and for all cases of practical relevance the Shannon entropy per internal random bit is in a vicinity of 1. Hence the Shannon entropy provides a sound estimate for the average guessing workload, which justifies the use of the Shannon entropy for physical RNGs in place of the more conservative min-entropy. (Min-entropy may be more suitable than the Shannon entropy for specific guessing problems that deal with very imbalanced probability distributions.) We point out in many cases it is much easier to compute the Shannon entropy than the min-entropy (cf. [24], Subsect. 5.2, for a more comprehensive treatment of this question).

2.2 Central Definitions and Goals

The core of a physical RNG is its *noise source* (noisy diode(s), oscillators etc.) Normally, the noise source produces a time-continuous analog signal that is (usually) digitized after uniform time intervals. The digitized values are called *das random numbers* where 'das' stands for *digital analog signal*. The das-random numbers may be algorithmically postprocessed, giving the so-called *internal random numbers*. Algorithmic postprocessing may increase the entropy per bit, but only at cost of performance (data compression). If the entropy of the das-random numbers is sufficiently large the algorithmic postprocessing may be saved in favour of better performance. Online and tot tests shall detect non-tolerable weaknesses while the RNG is in operation, e.g. caused by a total breakdown of the noise source. Upon external request the RNG outputs *external random numbers*.

The evaluation of physical RNGs falls into two parts. At first the generic design has to be considered. The central goal is to quantify (at least a lower bound for) the entropy per random bit. Unfortunately, entropy cannot be measured as voltage or temperature. Instead, entropy is a property of random variables and not of observed realizations. In particular, entropy cannot be ensured by passing a collection of statistical blackbox tests [21]. Note that usually even pseudo-random sequences pass these tests. To quantify entropy we have to study the distribution of the random numbers, or more precisely, the distribution of the underlying random variables.

Definition 1. *Random variables are denoted with capital letters.* Realizations *of these random variables, i.e. values that are assumed by these random variables, are denoted by the respective small letters. For instance, the das random numbers $r_1, r_2, \ldots$ are interpreted as realizations of random variables $R_1, R_2, \ldots$. We denote the internal random numbers and the underlying random variables by $y_1, y_2, \ldots$ and $Y_1, Y_2, \ldots$, respectively.*

External random numbers are not under control of the RNG designer. Since the external random numbers are usually concatenations of the internal random numbers it is yet natural to focus on

$$H(Y_{n+1} \mid Y_1 = y_1, \ldots, Y_n = y_n) \tag{2}$$

which corresponds to the real-life situation that an adversary knows a subsequence $y_1, y_2, \ldots, y_n$ of internal random numbers, e.g. due to openly transmitted challenges or session keys which the adversary received legitimately.

In the following the random variables $R_1, R_2, \ldots$ describe the stochastic behaviour of the das random numbers. Their distributions clearly depend on the noise source and the digitization mechanism. Usually, it is not feasible to determine these distributions exactly. At least in a strict sense the exact distribution depends on the characteristics of the components of the particular noise source, and these characteristics may differ to some extent even for RNGs from the same production series. A sound security evaluation of a physical RNG should be based on a *stochastic model*.

Stochastic Model. Ideally, the stochastic model comprises a *family of distributions* that contains the distribution of the internal random numbers. Since this is often too complicated in practice it suffices if the stochastic model defines a family of probability distributions that contains the distribution of the das-random numbers or even merely of 'auxiliary' random variables *provided that this is sufficient to attain our declared goal*, namely to verify a lower entropy bound for the internal random numbers. (This is the case in Sect. 4, for instance.)

Example 1. (Repeated tossing of a single coin) Since coins have no memory it is reasonable to assume that the random variables R_j are independent and binomially $B(1, p)$-distributed with $p \in [0, 1]$, defining a one-parameter family of probability distributions. Given a particular coin the parameter p may be guessed by tossing it a large number of times. The estimate $\widetilde{p}$ can be used to estimate the entropy. The entropy of the internal random numbers depends on p and the algorithmic postprocessing (if there is any).

For 'real life' RNGs formulation and analysis of the stochastic model is usually more complicated than in Example 1. For most RNG designs it is reasonable to assume that the sequence $R_1, R_2, \ldots$ is (strictly) stationary (i.e. time-invariant), at least within time periods that are large compared to the output rate. Drifts of process parameters within the life cycle of the RNG (e.g. due to ageing effects) are not problematic if the distribution remains in the acceptable part of the specified class. In a first step we are interested in

$$H(R_{n+1} \mid R_1 = r_1, \ldots, R_n = r_n) \tag{3}$$

for any history $r_1, \ldots, r_n$, or at least in the average conditional entropy

$$H(R_{n+1} \mid R_1, \ldots, R_n). \tag{4}$$

For dependent random variables the calculation of (4) is in general easier than (3). Due to tolerances of components, ageing effects or because of a total breakdown of the noise source a concrete RNG may output considerably weaker random numbers than the carefully investigated RNG prototypes. Online tests and tot tests ('total failure test') shall detect non-tolerable weaknesses while the RNG is in operation. There do not exist statistical tests that are universally strong for any RNG design. Instead, tests should be tailored to the stochastic model of the das random numbers. Depending on the conditions of use (strength of the device that contains the RNG, environmental conditions etc.) it may also be necessary that the design is also be resistant against active adversaries; at least physical sensors or the online test should detect induced weaknesses. The second task in a security evaluation is thus to verify the effectiveness of the online and tot tests and the consequence of noise alarms [20, 21, 25]. We will briefly address these aspects in Section 6.

Remark 2. A reasonable stochastic model is the core of any CC (Common Criteria) evaluation with regard to the evaluation guidance AIS 31 [2, 12], which has been effective in Germany since 2001. We point out that besides physical RNGs with cryptographic postprocessing the international ISO norm [10] also permits physical RNGs without cryptographic postprocessing if a sound stochastic model confirms that the random numbers have enough entropy and if effective online tests are applied.

3 An RNG based on Noisy Diodes

Figure 3 illustrates the principle of RNG which random source consists of two identical noisy diodes. E.g. Zener diodes have a reverse avalanched (depending on diode type 3 - 4 Volt or about 10 V) and produce more than 1mV noisy voltage on about 10 MHz. The outlets of both diodes are symmetrically input to an operational amplifier in order to amplify the difference of noise voltages. We first point out that a construction with only one noisy diode facilitates a manipulation by active adversaries, e.g. by external electromagnetic fields. The circuitry of AC coupling, the negative feedback for the operational amplifier, the stabilizing the power supply or compensating effects of temperature are not shown here. The operational amplifier output is provided to a Schmitt trigger where the mean voltage of the amplifier meets the threshold of the Schmitt trigger. The output signal of the Schmitt trigger consists of zeros and ones with random lengths.

Its 0-1 crossings clock an intermediate flip-flop that inverts the D-input of the last flip-flop which is latched by the clock. The das random number r_n is the number of 0-1-crossings within the n^{th} clock cycle. Consequently, $y_{n+1} = y_n \oplus r_{n+1} (\mathrm{mod}\, 2)$ where y_n and y_{n+1} denote the internal random numbers in Step n and $n+1$, respectively.

The quality of the das random numbers depends on the randomness of the number of these impulses between subsequents clock cycles. Unlike in a design

that exploits both, 0-1- and 1-0-crossings it is irrelevant whether the intervals of 0 and 1 are identically distributed. This makes the design more robust on cost of halving the output rate.

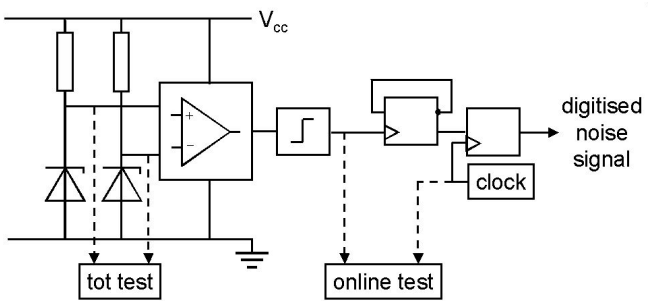

Fig. 1. RNG with two noisy diodes

For this RNG design only the average length betwen consecutive 0-1-crossings is relevant for the 0/1-ratio of the das random numbers. The distribution of this length may also change, causing a smaller or larger number of switchings within the particular clock cycles. If this number becomes too small this may reduce entropy. However, online and tot test should detect such a behaviour (cf. Subsect.6).

One part of the tot test checks separately the generation of noisy voltage for each diode. It detects breakdown or anormality of the noise. The online test checks statistically the quality of the digitised noise signal. Clearly, long-term aging effects may be neglected in that regard.

4 Formulation and Analysis of the Stochastic Model

In this section we formulate and analyze a stochastic model for the RNG design dfrom the last section. Interestingly, the same stochastic model fits also to to other RNG designs as well (cf. Remark 3(i)).

In the following we assume that the analog part of the noise source is in equilibrium state (because a sufficient amount of time has passed since the start of the system; a fraction of a second should suffice). We start with the analysis of the das random numbers $r_0, r_1, \ldots$ at time $t = 0$. The internal random numbers $y_1, y_2, \ldots$ are latched at equidistant times $s_1 := s, \ldots, s_j := js, \ldots$ where $s > 0$ denotes the cycle length of the clock that latches the final flip-flop (cf. Fig. 1 and Fig. 2). The Schmitt trigger switches at each 0-1 crossing of the voltage with the threshold value. The das random number r_n denotes the number of 0-1-switchings of the Schmitt trigger in the time interval $I_n := (s_{n-1}, s_n] = ((n-1)s, ns]$. Clearly

$$y_n \equiv y_{n-1} + r_{n+1} \equiv y_0 + r_1 + \cdots + r_n (\mathrm{mod} 2) \quad \text{for } n \geq 1 \tag{5}$$

where y_0 denotes the internal random number at time 0. Our goal is to determine a lower bound for

$$H(R_{n+1} \mid R_1, \ldots, R_n) \quad \text{and finally for} \quad H(Y_{n+1} \mid Y_0, Y_1, \ldots, Y_n), \tag{6}$$

the conditional entropy per das-random number, resp. the conditional entropy per internal random number. Note that the second formula corresponds to the real-world situation where an adversary knows several internal random numbers $y_0, y_1, y_2, \ldots, y_n$ (cf. Sect. 2). Since the algorithmic postprocessing is very elementary results on the das random numbers can directly be transferred to the internal random numbers.

Definition 2. *As usually, iid stands for 'independent and identically distributed'. A sequence of random variables $X_1, X_2, \ldots$ is called* stationary *if for each $r \geq 1$ the distribution of $(X_{m+1}, \ldots, X_{m+r})$ does not depend on the shift parameter m. The* generalized variance *of the sequence $X_1, X_2, \ldots$ is defined as*

$$\sigma^2 = \mathrm{Var}(X_1) + 2 \sum_{i=1}^{\infty} E\left((X_1 - \mu)(X_i - \mu)\right). \tag{7}$$

The sequence $X_1, X_2, \ldots$ is called q-dependent *if the vectors $(X_1, \ldots, X_a)$ and $(X_b, \ldots, X_n)$ are independent whenever $b - a > q$.*

As usually, $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. The cumulative distribution function of the standard normal distribution $N(0,1)$ is denoted with Φ, i.e. $\Phi(x) = \int_{\infty}^{x} e^{-t^2/2}\, dt/\sqrt{2\pi}$ for $x \in \mathbb{R}$.

Stochastic Model. We interpret the lengths $t_1, t_2, \ldots$ of the time intervals between consecutive 0-1-switchings as realizations of a q-dependent stationary stochastic process $T_1, T_2, \ldots$. We set $\mu := E(T_1)$ and $\sigma_T^2 := \mathrm{Var}(T_1)$ while the generalized variance of $T_1, T_2, \ldots$ simplifies to

$$\sigma^2 = \sigma_T^2 + 2 \sum_{i=1}^{q+1} E\left((T_1 - \mu)(T_i - \mu)\right) \tag{8}$$

We assume $\sigma_T^2 > 0$ (otherwise the das-random numbers were deterministic), $E(|T_j|^3) < \infty$ (needed for the proof of Lemma 1(iii); cd. also Remark 3) and $\mathrm{Prob}(T_1 = 0) = 0$.
The term z_n denotes the index of the first 0-1-switching that follows time $s_n = ns$ (i.e., when the clock latches the n^{th} time) while $w_n := t_{z_n} - s_n$. That is, w_n equals the time span from s_n to the next $0-1$-switching. In particular, $w_0 + t_1 + \cdots + t_{z_n - 1} \leq s_n < w_0 + t_1 + \cdots + t_{z_n}$. Recall that the stochastic model of an RNG shall allow to determine (a lower bound for) the conditional entropy $H(Y_{n+1} \mid Y_0, \ldots, Y_n)$. This defines our central goal.

More abstract, the random variables can be described as follows:

$$T_1, T_2, \ldots \quad \text{are stationary} \tag{9}$$

$$R_n := Z_n - Z_{n-1} \quad \text{for} \tag{10}$$

$$Z_n := min_{m \in N}\{W_0 + T_1 + T_2 + \ldots + T_m > s_n\} \tag{11}$$

Remark 3. (i) Relations (9) to (11) remain valid if we replace the two noisy diodes by a single diode.
(ii) We note that (9) to (11) also fits to the RNG design that was introduced in [23] and analyzed intensively in [22]. We note that the noise source consisted of two independent ring oscillators. In [22] we assumed $W_0 = 0$ for simplicity, which had little impact since the ratio s/μ (and thus the das random numbers $r_1, r_2, \ldots$) were very large.
(iii) The assumption that the T_j are q-dependent may be relaxed as long as the central limit theorem (for dependent random variables) remains valid (cf. (16), for instance).

For these reasons it is worthwhile to study the system (9) to (11) under general (weak) assumptions and for specific conditions on the distribution of the T_j (e.g., for iid or Markovian T_j). Note, however, that even if the das random numbers of different RNGs can be modelled by the system (9) to (11) the distributions of the random variables $T_1, T_2, \ldots$ and, consequently, of $R_1, R_2, \ldots$ and $Y_1, Y_2, \ldots$ may be very different.

Remark 4. (i) Due to the nature of shot noise one may assume that q is very small, presumably $q \leq 1$ (cf. Sect. 5 and Remark 3(iv)).
(ii) In our context $\mu \ll s$ so that at least one 0-1-switching occurs in each time interval with overwhelming probability. Then z_n equals the index of the first 0-1-switching in the interval $(sn, s(n+1)]$.

Assumption. Lemma 2 (Appendix) considers the 'transfer' of the strict stationarity property. Unlike $\mathrm{Prob}(T'_j \geq s)$ the probability $\mathrm{Prob}(T_j \geq s)$ is not exactly 0 but negligible (since $\mu \ll s$), and it is reasonable to assume that for 'large' index j the term $T_1 + \cdots + T_j (\mathrm{mod}\, s)$ is uniformly distributed on $[0, s)$ ($\rightarrow$ uniformity assumption on S'_j). Note that the periods between $0-1$ switchings from the start of the RNG to time 0 can be described by random variables T_j with negative indices. The assumptions on the T_j seem to be natural and very mild. With regard to Lemma 2 we may assume in the following that to the random variables $(W_j, R_j), R_j, W_j, Y_j$ and Z_j are stationary.

Definition 3. *The cumulative distribution functions of the random variables T_j and W_n are denoted by $G_T(\cdot)$ and $G_W(\cdot)$. For $u \in (0, \infty)$ the random variable $V_{(u)} := \inf\left\{\tau \in \mathbb{N} \mid \sum_{j=1}^{\tau+1} T_j > u\right\} = \sup\left\{\tau \in \mathbb{N} \mid \sum_{j=1}^{\tau} T_j \leq u\right\}$ quantifies the number of 0-1-switchings in the interval $[0, u]$.*

Lemma 1 collects some useful properties that will be needed later. Note that (12) formally confirms our intuition that knowing more random numbers cannot weaken the adversary's position. We point out that (12) may become false without the stationarity property, namely when R_n is easier to guess than R_{n+1}.

Lemma 1.

(i)
$$H(R_n \mid R_0, R_1, \ldots, R_{n-1}) \geq H(R_{n+1} \mid R_0, R_1, \ldots, R_n) \quad \textit{and} \qquad (12)$$
$$H(Y_n \mid Y_0, Y_1, \ldots, Y_{n-1}) \geq H(Y_{n+1} \mid Y_0, Y_1, \ldots, Y_n) \quad \textit{for all } n \in \mathbb{N}\,.$$

In particular, $\lim_{n\to\infty} H(R_{n+1} \mid R_1, \ldots, R_n)$ *and* $\lim_{n\to\infty} H(Y_{n+1} \mid Y_1, \ldots, Y_n)$ *exist.*
(ii) For $k \geq 1$ *we have*

$$\text{Prob}(V_{(u)} = k) = \text{Prob}\,(T_1 + \cdots + T_k \leq u) - \text{Prob}\,(T_1 + \cdots + T_{k+1} \leq u)\,. \quad (13)$$

Further,

$$\text{Prob}(V_{(u)} = 0) = 1 - \text{Prob}\,(T_1 \leq u)\,, \;\; \text{Prob}(V_{(u)} = \infty) = 0 \quad \textit{and} \quad (14)$$
$$H(V_{(u)}) < \infty. \quad (15)$$

(iii) The distributions of the random variables $(\sum_{j=1}^{k} T_j - k\mu)/(\sqrt{k}\sigma)$ *tend to the standard normal distribution as* k *tends to infinity. In particular,*

$$\text{Prob}\left(\frac{T_1 + \cdots + T_k - k\mu}{\sqrt{k}\sigma} \leq x\right) \longrightarrow_{k\to\infty} \Phi(x). \quad (16)$$

for each $x \in \mathbb{R}$.
If the random variables $T_1, T_2, \ldots$ *are iid the condition* $E(|T_j|^3) < \infty$ *may be dropped, and in particular* $\sigma^2 = \sigma_T^2$
(iv) Let $u = v\mu$ *with* $v >> 1$. *Then*

$$\text{Prob}\left(V_{(v\mu)} = k\right) \approx \Phi\left(\frac{v-k}{\sqrt{k}} \cdot \frac{\mu}{\sigma}\right) - \Phi\left(\frac{v-(k+1)}{\sqrt{k+1}} \cdot \frac{\mu}{\sigma}\right) \quad \textit{for } k \geq 1 \quad (17)$$
$$\text{Prob}\left(V_{(v\mu)} = 0\right) \approx 1 - \Phi\left((v-1)\frac{\mu}{\sigma}\right). \quad (18)$$

The distribution of the random variable $V_{(v\mu)}$ *(or more precisely, its approximation given by (17) and (18)) depends only on the ratios* μ/σ *and* $u/\mu = v$ *but not on the absolute values of the parameters* $\mu, \sigma^2, u = v\mu$. *The mass of* $V_{(v\mu)}$ *is essentially concentrated on those* k*'s with* $k \approx v$. *Unless* k *is very small the interval*

$$J_k := \left[\frac{v-(k+1)}{\sqrt{k+1}} \cdot \frac{\mu}{\sigma}, \frac{v-k}{\sqrt{k}} \cdot \frac{\mu}{\sigma}\right) \quad \textit{has length } \approx \frac{\mu}{\sigma} \cdot \frac{v+k}{2k^{3/2}} \quad (19)$$

(v) (iid case) If the random variables $T_1, T_2, \ldots$ *are iid then*

$$\lim_{n\to\infty} \text{Prob}(W_n \leq x) = \frac{1}{\mu}\int_0^x (1 - G_T(u))\, du =: G_W(x). \quad (20)$$

for any distribution of W_1. *If* $(W_n)_{n\in N}$ *is stationary* '$\lim$' *may be omitted. If* $G_T(\cdot)$ *is continuous (or equivalently, if* $\text{Prob}(T_1 = y) = 0$ *for all* $y \in [0, \infty)$*) then* $G_W(\cdot)$ *has density* $g(x) := (1 - G_T(x))/\mu$.

Proof. By Assumption 4 the random variables R_j and Y_j are stationary. Hence, (e.g.)

$$H(Y_n \mid Y_1, \ldots, Y_{n-1}) = H(Y_{n+1} \mid Y_2, \ldots, Y_n) \geq H(Y_{n+1} \mid Y_1, \ldots, Y_n),$$

and since entropy is non-negative this verifies (i). Assertions (ii), (iii) and the first assertions of (iv) follow from Lemma 1 and Lemma 2(ii) in [22]. We merely mention that (ii) applies a version of the Central Limit Theorem for dependent random variables that was proved in [11]. The remaining assertions in (iv) demand elementary but careful computations. (Note that $(\sqrt{k+1}-\sqrt{k})(\sqrt{k+1}+\sqrt{k}) = 1$ and $\sqrt{k} \approx \sqrt{k+1}$.) Formula (20) was verified in [9] (4.10), and the second assertion of (v) follows by differentiation.

Under mild regularity assumptions on the $T_1, T_2, \ldots$ plausible heuristic arguments indicate that

$$H(Y_{n+1} \mid Y_1, \ldots, Y_n) \geq \min\{H(V_{(s-u)}(\mathrm{mod}2) \mid u \in [0, \mu + a\sigma)\} G_W(\mu + a\sigma). \tag{21}$$

for moderate parameter $a > 0$. We point out that for $n = 0$ or if the T_j are iid (21) is valid for any positive a. Due to the lack of space we omit details. Theorem 1 collects central results. Theorem 1 focuses on the entropy of the internal random numbers. Cancelling the term '$(\mathrm{mod}\, 2)$' in (24), (21), (25) and (26) yields entropy estimates for the das random numbers. Equation (29) can be used to compute the autocovariance function and the autocorrelation function of the random variables $R_1, R_2, \ldots$ (see Example 2).

Theorem 1.

$$(i) \qquad \mathrm{Prob}(R_{n+1} = k) \approx \int_0^s \mathrm{Prob}(V_{(s-u)} = k-1)\, G_W(du) \qquad \textit{and} \tag{22}$$

$$\mathrm{Prob}(Y_{n+1} = k) \approx \int_0^s \mathrm{Prob}(V_{(s-u)} \equiv k-1(\mathrm{mod}\, 2))\, G_W(du) \tag{23}$$

$$H(Y_{n+1}) \geq H(Y_{n+1} \mid W_n) \approx \int_0^s H(V_{(s-u)}(\mathrm{mod}\, 2))\, G_W(du) \tag{24}$$

with equality for iid random variables T_j.
(ii) Substituting the integrands in (22) to (24) by $\mathrm{Prob}(V_{(s-u)} = k-1 \mid W_0 = u)$, $\mathrm{Prob}(V_{(s-u)} \equiv k-1(\mathrm{mod}\, 2) \mid W_0 = u)$, *and* $H(V_{(s-u)}(\mathrm{mod}\, 2) \mid W_0 = u)$, *resp., provides equality also for the general case. Note that for dependent T_j the conditional terms implicitly define conditions on the random variables $T_1, T_2, \ldots$ and thus on $V_{(s-u)}$.*
(iii) (iid case) If the sequence $T_1, T_2, \ldots$ is iid

$$H(Y_{n+1} \mid Y_0, \ldots, Y_n) \geq \int_0^s H(V_{s-u)}(\mathrm{mod}\, 2))\, G_W(du) \;\; \textit{for all } n \in \mathbb{N}\,. \tag{25}$$

If $G_T(\cdot)$ is continuous the right-hand side of (25) reads

$$\int_0^s H(V_{(s-u)}(\mathrm{mod}\, 2))\, \frac{1}{\mu}(1 - G_T(u))\, du. \tag{26}$$

$$(iv) E((R_1 + \cdots + R_j)^k) = \int_0^{js} E((V_{(js-u)} + 1)^k \mid W_0 = u))\, G_W(du) \tag{27}$$

$$\approx \int_0^{js} E((V_{(js-u)} + 1)^k)\, G_W(du) \;\; \textit{for each } k \in \mathbb{N} \tag{28}$$

with equality for iid random variables T_j. The stationarity of the R_j implies

$$E((R_1 + \ldots + R_j)^2) = jE(R_1^2) + 2\sum_{i=2}^{j}(j+1-i)E(R_1R_i) \tag{29}$$

Proof. By stationarity $R_{n+1} \mid W_n = u$ is distributed as $V_{(s-w_n)} + 1 \mid W_0 = u$, and thus $Y_{n+1} \mid W_n = u$ as $V_{(s-w_n)} + 1 (\mathrm{mod}\, 2) \mid W_0 = u$. Formulae (22) to (24) and (ii) follow immediately from the stationarity of the random variables $R_1, R_2, \ldots$, $W_1, W_2, \ldots$ and $Y_1, Y_2, \ldots$, respectively. Within this proof ν_n and $\nu_{n|y_0,\ldots,y_n}$ denotes the distribution of W_n, resp. of the conditional random variable $(W_n \mid Y_0 = y_0, \ldots, Y_n = y_n)$. In this notation

$$\begin{aligned} &H(Y_{n+1} \mid Y_0 = y_0, \ldots, Y_n = y_n) \geq H(Y_{n+1} \mid Y_0 = y_0, \ldots, Y_n = y_n, W_n) \qquad (30) \\ &= \int_0^\infty H(Y_{n+1} \mid Y_j = y_j, j \leq n; W_n = u)\, \nu_{n|y_0,\ldots,y_n}(du) \end{aligned}$$

If the T_j are iid for all $n \in N$ the vector $(T_{z_n+1}, T_{z_{n+1}+2}, \ldots)$ is distributed as $(T_1, T_2, \ldots)$, regardless of u and the history $y_0, \ldots, y_n$. In particular, the integrand of the right-hand side of (30) only depends on u. More precisely, for any $y_0, \ldots, y_n$ it equals $H(V_{(s-u)} + 1 (\mathrm{mod}\, 2)) = H(V_{(s-u)}) (\mathrm{mod}\, 2)$. Altogether

$$\begin{aligned} &H(Y_{n+1} \mid Y_0, \ldots, Y_n, W_n) \\ &= \sum_{y_0,\ldots,y_n \in \{0,1\}} \mathrm{Prob}(Y_0 = y_0, \ldots, Y_n = y_n) \int_0^\infty H(V_{(s-u)} (\mathrm{mod}\, 2))\, \nu_{n|y_1,\ldots,y_n}(du) \\ &= \int_0^\infty H(V_{(s-u)} (\mathrm{mod}\, 2))\, \nu_n(du) = \int_0^s H(V_{(s-u)} (\mathrm{mod}\, 2))\, G_W(du) \end{aligned}$$

in the iid case. The final equation follows from the fact that $\mathrm{Prob}(W_{n+1} > s) = 1 - G_W(s) \approx 0$ since $s >> \mu$. This proves (25), and (26) follows immediately from Lemma 1(v). The sum $R_1 + \cdots + R_n = Z_{n+1} - Z_0$ is distributed as $V_{(ns-W_0)} + 1$, which proves (27). For iid T_j the history (expressed by W_0) is irrelevant, which implies (28). The stationarity of the T_j finally implies (29).

Remark 5. (Robustness) Formulae (24), (25) and (26) provide entropy estimators that are robust against moderate deviations of the distribution of $T_1, T_2, \ldots$. By (17) and (18) the term $H(V_{(s-u)})$ depends only on μ and σ. In (26) the term $1 - G_T(\cdot)$ is monotonically decreasing, which additionally supports robustness.

Example 2. Assume that the random variables T_j are iid Erlang-$(2/\mu, 2)$-distributed. (This is the case, for instance, if the intermediate times between (0-1)- and (1-0)-crossings as well as the intermediate time between (1-0)- and (0-1)-crossings are independent and identically exponentially distributed with parameter $\lambda = \mu/2$; cf. also Sect. 5.) As usually, $\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y)$ denotes the covariance of the random variables X and Y, while gen.$\mathrm{Var}(R_1, \ldots)$ stands for the generalized variance of the sequence $R_1, R_2, \ldots$.

	$s = 6.0\mu$	$s = 6.5\mu$	$s = 7.0\mu$	$s = 9.0\mu$	$s = 10.5\mu$
$E(R_1)$	6.000	6.500	7.000	9.000	10.500
$\mathrm{Var}(R_1)$	3.458	3.709	3.959	4.959	5.709
$\mathrm{cov}(R_1, R_2)$	-0.2299	-0.2295	-0.2291	-0.2289	-0.2287
$\mathrm{cov}(R_1, R_3)$	-0.0012	-0.0006	-0.0004	0.0003	0.00004
$\mathrm{gen.Var}(R_1, \ldots)$	2.990	3.246	3.499	4.502	5.254
$H(R_{n+1} \mid R_1, \ldots, R_n)$	2.750	2.815	2.875	3.073	3.193
$H(Y_{n+1} \mid Y_1, \ldots, Y_n)$	0.9998	0.99992	0.99996	0.999994	0.999994

Table 1. Numerical Results

The entries from Table 1 were computed with Theorem 1; we used (28), (29) and (26). We point out that inequality (21) provides more pessimistic lower entropy bounds for $H(Y_{n+1} \mid Y_1, \ldots, Y_n)$, namely 0.333, 0.333, 0.333, 0.801, and 0.988. Compared to $\mathrm{Var}(R_1)$ and $\mathrm{cov}(R_1, R_2)$ the covariances $\mathrm{cov}(R_n, R_{n+k})$ are negligible for $k \geq 2$.

5 Practical Experiments

As pointed out in Remark 3 relations (9) to (11) fit to various RNG designs. The distribution of the random variables T_j and thus of R_j and Y_j yet depend on the design but also on the concrete implementation. We performed measurements for a prototype [3] of a particular physical RNG (cf. Fig. 2 that uses the generic design discussed in Sect. 3. Large measurement series verified that both the intermediate times between (0-1)- and (1-0)-crossings as well as the intermediate times between (1-0)- and (0-1)-crossings identically exponentially distributed. Figure 2 shows the relative frequencies of period lengths between consecutive 0-1-crossings. For the statistical calculations the statistical environment "R" (cf. www.r-project.org) was used.

The average period length is ≈ 128.9 ns so that the RNG could output 700 kBit random numbers if $s = 10.5 \cdot 128.9$ ns, for instance. Maximum-Likelihood tests indicate that the one-dimensional empirical distribution from Figure 2 can be approximated by an Erlang$(0.0155/ns, 2)$-distribution. Contingency tests did not contradict the hypothesis that T_j and T_{j+1} are independent (97 from 99 tests with significance level 0.01 were passed). Thus we used Example 2 (cf. Table1) to conclude that $H(Y_{n+1} \mid Y_1, \ldots, Y_n) > 1 - 10^{-5}$ for $s = 10.5 \cdot 128.9$ ns. We mention that this conclusion is supported by Remark 5 and the fact that for the (exact) Erlang distribution even considerably smaller s suffices to guarantee this bound.

We point out that Theorem 1 can also directly be applied to empirical data. This is in particular relevant if the empirical distribution cannot be approximated by any 'well-known' distribution. Of course, an approximator for $G_T(\cdot)$ follows immediately from the empirical distribution of the lengths t_j. If the T_j are iid

[3] courteously provided by Mr. Bergmann, www.ibbergmann.org

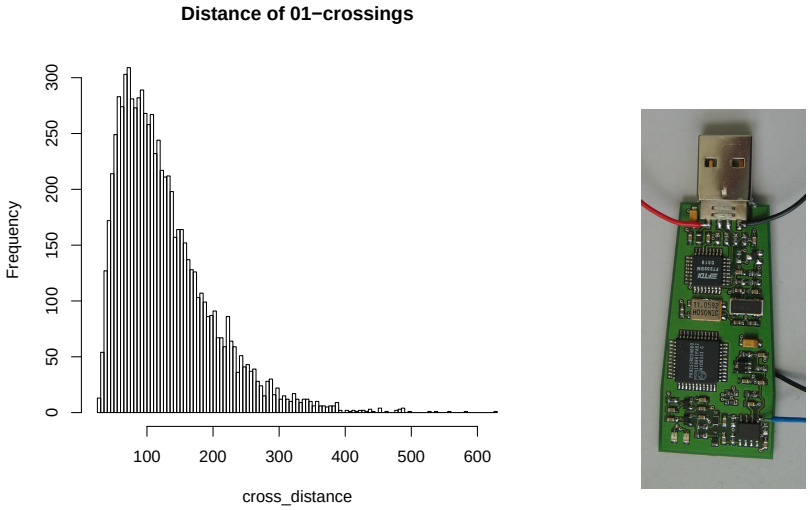

Fig. 2. Empirical distribution of distances between 0-1 crossings / picture of the RNG

with density this approximator can be substituted into (26). For arbitrary T_j to given clock cycle length s the cumulative distribution function G_W may be estimated from traces $t_1, t_2, \ldots$. We are going to implement such a routine in "R", which shall enables us to verify the entropy bound $1 - 10^{-5}$ with an alternative approach.

6 Online Tests

The conditional entropy $H(Y_{n+1} \mid Y_0, \ldots, Y_n)$ is closely related to the entropy of the random variables $V_{(s-u)}$ (cf. Theorem 1 and (21)). By (17) and (18)the entropy $H(V_{(s-u)})$ depends only on the ratios μ/σ and $(s-u)/\mu$. Moreover, unless the cycle length s of the external clock is small compared to μ arguments u with $s - u \approx s$) provide the esential contribution to the integral. Since s is assumed to be fixed it is natural (and effective) to estimate the process parameters $\mu = E(T_j)$ and the generalized variance σ^2 of $T_1, T_2, \ldots$ (cf. (7)) while the RNG is in operation. Unfortunately, this required an internal clock with high resolution, which may be too costly for many applications.

For most applications it is more convenient to check the process parameters μ and σ^2 indirectly, namely by estimating the mean value $\mu_R := E(R_j)$ and the generalized variance of the stationary sequence $R_1, R_2, \ldots$. In a first step, appropriate intervals I_μ and I_{σ^2} for the process parameters μ and σ^2 have to be specified. Applying Theorem 1(iv) yields intervals I_{μ_R} and $I_{\sigma_R^2}$ that contain μ_R and σ_R^2 if μ and σ^2 are contained in I_μ and I_{σ^2}. For a physical RNG that meets

(9) to (11) we propose the following generic test procedure. Input data for the statistical tests are das random numbers $r_1, \ldots, r_{m+M}$.

- Compute $\mathrm{av}(r_1, \ldots, r_m) := (r_1 + \cdots + r_m)/m$.
  - (tot test) If $\mathrm{av}(r_1, \ldots, r_m) < a \Rightarrow$ noise alarm
  - (online test) If $\mathrm{av}(r_1, \ldots, r_m) \notin [b_1, b_2] \Rightarrow$ noise alarm
- (online test) Use das random numbers $r_{m+1}, \ldots, r_{m+M}$ to decide whether $\sigma_R^2 \in [c_1, c_2]$. If the decision is negative: $\Rightarrow$ noise alarm

Due to the limited task of a tot test the parameter a may be selected very small to prevent false noise alarms. Alternatively, this task may be covered by a physical sensor. The parameters b_1, b_2, c_1, c_2 must fulfil $I_{\mu_R} \subseteq [b_1, b_2]$ and $I_{\sigma_R^2} \subseteq [c_1, c_2]$. As usual, the selection of these parameters depends on the application scheme of the online test (continuously, on demand etc.).

The distribution of $\mathrm{av}(R_1, \ldots, R_m)$ can be computed by (22) with upper integration boundary ms in place of s. Alternatively, the central limit theorem (which holds under weak assumptions on the R_j) might be applied to the random variablers R_j since μ_R and σ_R^2 can be computed with (29). The second online test should be tailored to the distribution of the random variables R_j for the concrete RNG. The generalized variance σ_R^2 can be estimated directly, or the relevant set of covariances $\mathrm{cov}(R_n, R_{n+k})$ may be estimated. A precise computation of the failure probability for $[c_1, c_2]$ is more complicated than for the arithmetic mean. (Selecting large sample size M may simplify this problem.) Typically, joint moments (e.g. $E(R_{j-1}R_j^2R_{j+1})$) have to be computed. This may be done on basis of theoretical considerations or on measurement series, or by stochastic simulations (with pseudorandom numbers $\widetilde{t}_1, \widetilde{t}_2, \ldots$ that are generated with regard to the specified distribution of the random variables T_j).

For the RNG considered in Section 5 (Erlang distribution) σ^2 and thus μ_R and σ_R^2 are functions of μ. If influences from possible fault attacks will either reliably be detected by physical sensors or prevented by the implementtation, the second online test might be dropped.

7 Conclusions

We discussed technical aspects for an RNG design that uses two noisy diodes in place of one, and we explained the advantages. We formulated and analyzed a general stochastic model that fits to this and to other RNG designs. We proved a theorem that provides robust estimators for a lower entropy bound for the generated random random numbers. We applied our results to a particular physical RNG and estimated its entropy. Finally, we discussed a generic design for online tests.

References

1. AIS 20: Functionality Classes and Evaluation Methodology for Deterministic Random Number Generators. Version 1 (02.12.1999) (mandatory

if a German IT security certificate is applied for; English translation). www.bsi.bund.de/zertifiz/zert/interpr/ais20e.pdf
2. AIS 31: Functionality Classes and Evaluation Methodology for Physical Random Number Generators. Version 1 (25.09.2001) (mandatory if a German IT security certificate is applied for; English translation). www.bsi.bund.de/zertifiz/zert/interpr/ais31e.pdf
3. ANSI X9.82, Random Number Generation (Draft Version).
4. P. Billingsley: Probability and Measure. Wiley, New York 1979.
5. M. Bucci, L. Germani, R. Luzzi, A. Trifiletti, M. Varanonuovo: A High-Speed Oscillator-Based Truly Random Number Source for Cryptographic Applications, IEEE Trans. Computers, Vol. 52, 2003, 403–409.
6. M. Bucci, R. Lucci: Design of Testable Random Bit Generators. In: J. Rao, B. Sunar (eds.): Cryptographic Hardware and Embedded Systems — CHES 2005. Springer, Lecture Notes in Computer Science, Vol. 3659, Berlin (2005), 147–156.
7. H. Bock, M. Bucci, R. Luzzi: An Offset-Compensated Oscillator-Based Random Bit Source for Security Applications. In: M. Joye, J.-J. Quisquater (eds.): Cryptographic Hardware and Embedded Systems — CHES 2004. Springer, Lecture Notes in Computer Science, Vol. 3156, Berlin (2004), 268–281.
8. M. Dichtl, J. Golic: High-Speed True Random Number Generation with Logic Gates Only. In: P. Paillier, I. Verbauwhede (eds.): Cryptographic Hardware and Embedded Systems — CHES 2007, Springer, Lecture Notes in Computer Science 4727, Berlin 2007, 45–62.
9. W. Feller: An Introduction to Probability Theory and Its Application (Vol. 2). Wiley, New York 1965.
10. ISO / IEC 18031 'Random Bit Generation'. November 2005.
11. W. Hoeffding, H. Robbins: The Central Limit Theorem for Dependent Random Variables. Duke Math. J. 15 (1948), 773–780.
12. W. Killmann, W. Schindler: A Proposal for: Functionality Classes and Evaluation Methodology for True (Physical) Random Number Generators. Version 3.1 (25.09.2001), mathematical-technical reference of [2] (English translation); www.bsi.bund.de/zertifiz/zert/interpr/trngk31e.pdf
13. D.P. Maher, R.J. Rance: Random Number Generators Founded on Signal and Information Theory. In: Ç.K. Koç, C. Paar (eds.): Cryptographic Hardware and Embedded Systems — CHES 1999. Springer, Lecture Notes in Computer Science, Vol. 1717, Berlin (1999), 219–230.
14. S. Mandal, S. Banerjee: An Integrated CMOS Chaos Generator. In: S. Banerjee (ed.): 1^{st} Indian National Conference on Nonlinear Systems & Dynamics — NCNSD 2003. Kharagpur (India), (2003), 313–316.
15. G. Marsaglia: Diehard (Test Suite for Random Number Generators). www.stat.fsu.edu/˜geo/diehard.html
16. U. Maurer: A Universal Statistical Test for Random Bit Generators. J. Crypt. 5 (1992), 89–105.
17. A. Rukhin et al.: A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. NIST Special Publication 800–22 with revisions dated 15.05.2001. csrc.nist.gov/rng/SP800-22b.pdf
18. J.O. Pliam: The Disparity Between the Work and the Entropy in Cryptology (01.02.1999). eprint.iacr.org/complete/
19. W. Schindler: Functionality Classes and Evaluation Methodology for Deterministic Random Number Generators. Version 2.0 (02.12.1999), mathematical-technical reference of [1] (English translation); www.bsi.bund.de/zertifiz/zert/interpr/ais20e.pdf

20. W. Schindler: Efficient Online Tests for True Random Number Generators. In: Ç.K. Koç, D. Naccache, C. Paar (eds.): Cryptographic Hardware and Embedded Systems — CHES 2001. Springer, Lecture Notes in Computer Science, Vol. 2162, Berlin (2001), 103–117.
21. W. Schindler, W. Killmann: Evaluation Criteria for True (Physical) Random Number Generators Used in Cryptographic Applications. In: B.S. Kaliski Jr., Ç.K. Koç, C. Paar (eds.): Cryptographic Hardware and Embedded Systems — CHES 2002, Springer, Lecture Notes in Computer Science 2523, Berlin 2003, 431-449.
22. W. Schindler: A Stochastical Model and Its Analysis for a Physical Random Number Generator Presented at CHES 2002. In: K.G. Paterson (ed.): Cryptography and Coding — IMA 2003, Springer, Lecture Notes in Computer Science 2898, Berlin 2003, 276–289.
23. T. Tkacik: A Hardware Random Number Generator. In: B.S. Kaliski Jr., Ç.K. Koç, C. Paar (eds.): Cryptographic Hardware and Embedded Systems — CHES 2002, Springer, Lecture Notes in Computer Science 2523, Berlin 2003, 450–453.
24. XXX: Random Number Generators for Cryptographic Applications. To appear in: Ç.K. Koç (ed.): Cryptographic Engineering. Springer, Signals and Communication Theory, Berlin.
25. XXX: Evaluation Criteria for Physical Random Number Generators. To appear in: Ç.K. Koç (ed.): Cryptographic Engineering. Springer, Signals and Communication Theory, Berlin.

Appendix

Lemma 2. *(Stationarity Lemma) Let* $\ldots, T'_{-1}, T'_0, T'_1, \ldots$ *denote a doubly infinite sequence of stationary random variables with* $\mathrm{Prob}(T'_j \in [0,s)) = 1$ *and* $\mathrm{Prob}(T'_j = 0) < 1$. *Assume that the sequence* $\ldots, S'_{-1}, S'_0, S'_1, \ldots$ *fulfils* $S'_{j+1} - S'_j \equiv T'_{j+1} (\mathrm{mod}\, s)$ *for each integer* j. *Assume further that* S'_J *is uniformly distributed on* $[0,s)$ *and independent from the random variables* $\ldots, T'_{-1}, T'_0, T'_1, \ldots$ *for a particular integer* J.
(i) S'_j *is uniformly distributed on* $[0,s)$ *for each integer* j, *and the random variables* $\ldots, S'_{-1}, S'_0, S'_1, \ldots$ *are stationary.*
(ii) For $j \geq 1$ *let* z'_j *denote the* j^{th} *index* $m > 0$ *for which* $S'_m < S'_{m-1}$, *and* $W'_j := S'_{z_j}$. *For* $R'_j = Z'_j - Z'_{j-1}$ *the random vectors* (W'_j, R'_j) *and the random variables* W'_j, R'_j *and* $Y' := f(R'_j)$ *(with* $f\colon R \to \mathbb{R}$*) are stationary.*

Proof. For $k \geq 0$ trivially $S'_{J+k} \equiv S'_J + T'_{J+1} + \cdots + T'_{J+k} (\mathrm{mod}\, s)$, and the independence of S'_J and $T'_{J+1} + \cdots + T'_{J+k}$ proves the first assertion of (i). The case $k < 0$ can be handled analogously. We point out that the sequence $(S'_j - S'_{j-1}) (\mathrm{mod}\, s) \equiv T'_j (\mathrm{mod}\, s)$ is stationary. We claim that S'_{J+j} and $(T'_i, \ldots, T'_k)$ are independent for any triple of integers (j,i,k) with $i \leq k$. Let $M := \{J+1, \ldots, J+j, i, \ldots, k\}$ and assume $j \geq 0$ for the moment. Then $\mathrm{Prob}(S'_{J+j} \in A \mid T'_\tau = t'_\tau \text{ for all } \tau \in M\} = \mathrm{Prob}(S'_J + T'_{J+1} + \cdots + T'_{J+j} (\mathrm{mod}\, s) \in A \mid T'_\tau = t'_\tau \text{ for all } \tau \in M\} = \mathrm{Prob}(S_J \in (A - t'_{J+1} - \cdots - t'_{J+j}) (\mathrm{mod}\, s)) = \mathrm{Prob}(S'_J \in A)$ for each Borel subset $A \subseteq [0,s)$ and any realizations $t'_{J+1}, \ldots, t'_{J+j}, t'_i, \ldots, t'_k$ since the random variables S'_J and $(T'_{J+1}, \ldots, T'_{J+j}, T'_i, \ldots, T'_k)$ are independent, and S_J is uniformly distributed on $[0,s)$. This proves the claim for $j \geq 0$. For

$j \le 0$ we have $S'_J \equiv S'_{J+j} + T_{J+j+1} + \cdots + T'_J (\mathrm{mod}\, s)$, and the claim can be shown analogously. Let k and j be fixed for the moment. By the preceding $\mathrm{Prob}(S'_{j+1}, (T'_{j+2}, \ldots, T'_{j+k}) \in A \times B) = \mathrm{Prob}(S'_{j+1} \in A)\mathrm{Prob}(T'_{j+2}, \ldots, T'_{j+k} \in B) = \mathrm{Prob}(S'_1 \in A)\mathrm{Prob}(T'_2, \ldots, T'_k \in B) = \mathrm{Prob}(S'_1, (T'_2, \ldots, T'_k) \in A \times B)$ for any Borel subsets $A \subseteq [0, s)$ and $B \subseteq [0, s)^{k-1}$. Hence $(S'_1, T'_2, \ldots, T'_k)$ and $(S'_{j+1}, T'_{j+2}, \ldots, T'_{j+k})$ are identically distributed. Let the diffeomeorphism $\chi_k \colon [0, s)^k \to [0, s)^k$ be given by $\chi(x_1, \ldots, x_k) := (x_1, x_1 + x_2 (\mathrm{mod}\, s), \ldots, x_1 + \cdots + x_k (\mathrm{mod}\, s))$. Since $(S'_{j+1}, S'_{j+2}, \ldots, S'_{j+k}) = \chi_k(S'_{j+1}, T'_{j+2}, \ldots, T'_{j+k})$, the random vectors $(S'_1, S'_2, \ldots, S'_k)$ and $(S'_{j+1}, S'_{j+2}, \ldots, S'_{j+k})$ are identically distributed. Since j and k were arbitrary, this completes the proof of (i).

Let $j_1 > 0$ denote the smallest index for which $S'_{j_1} < S'_{j_1 - 1}$. Divide the random variables $\ldots, S'_{-1}, S'_0, S'_1, \ldots$ into increasing subsequences $\ldots, (\ldots, S'_{j_1 - 1})$, $(S'_{j_1}, \ldots, S'_{j_2 - 1}), (S'_{j_2}, \ldots), \ldots$ such that $S'_{j_m - 1} > S'_{j_m}$. (As $\mathrm{Prob}(T'_j = 0) < 1$ these subsequences are finite with probability 1.) Alternatively, these subsequences can be described by the sequence $(W'_j, R'_j)_{j \in Z}$ (and index j_0). For any $k \ge 1$, integers $r_1, \ldots, r_k \ge 1$ and subsets $A_1, \ldots, A_k \subseteq [0, s)$ the probability $\mathrm{Prob}((W'_{1+\tau}, R'_{1+\tau}) \in A_1 \times \{r_1\}, \ldots, (W'_{k+\tau}, R'_{k+\tau}) \in A_k \times \{r_k\})$ depends on the distribution of the $r := (r_1 + \cdots + r_k + 2)$-tuple $(S'_{j_1 - 1}, \ldots, S'_{j_1 + r - 2})$. Since the sequence $\ldots, (S'_1, \ldots, S'_r), (S'_2, \ldots, S'_{r+1}), \ldots$ is stationary the above probability is independent of τ. This proves the stationarity of $(W'_j, R'_j)_{j \in Z}$. The random variables W'_j, R'_j and Y'_j are functions of (W'_j, R'_j), which completes the proof of (ii).