Embedded evaluation of the statistical parameters of the thermal noise for online test of TRNG

Olivier Batard, David Lubicz

Embedded evaluation of the statistical parameters of the thermal noise for online test of TRNG

Olivier Batard, David Lubicz

November 6, 2023

Abstract

One of the most widely used framework to implement True Random Number Generators (TRNG) on an electronic ship is based on free running oscillators. In order to make sure that a TRNG has a given entropy rate, it is desirable to embed with it an online test of entropy. According to [1], the only way to do achieve that, is to use a statistical model of the TRNG and to have an online measurement method of the physical parameters of the model. In the context of oscillator based TRNG a method is proposed in [2] to assess the statistical parameters of the model proposed in [3]. In this paper, we explain that this method has some shortcomings, that the choice of its parameters, not clearly explained in [2] can have dramatic effects on its precision and reliability. In this paper, we propose a recipe for choosing good parameters and present some algorithmic tweaks to improve its robustness. We have extensively tested the new method with simulations and implemented it in hardware to verify its efficiency and practicality.

Keywords:

Hardware random number generators, free-running oscillators, stochastic models, entropy, dedicated statistical tests.

1 Introduction

Random numbers generators are essential security component used in every cryptographic hardware implementation for the production of confidential keys, initialization vectors, padding values, challenge for authentication protocols and also as random masks in side-channel attack countermeasures. They are usually composed of two stages: a non deterministic stage, called True Random Number Generator (TRNG) and a deterministic one, called Deterministic Random Number Generator (DRNG). The TRNG serves as an entropy source to guarantee unpredictability of generated numbers against an adversary with an unlimited computational power. The DRNG, which uses TRNG output as a seed, ensures that the output of the random number generator is unpredictable for an attacker with a bounded computational power even if the TRNG partially or temporarily fails according a security model such as the one of [].

This paper deals with the TRNG stage of a random number generator. It is still a challenging task to design such a device with a high level of certainty on its security which is measured for cryptographic application by its entropy rate. In order to do so, according to [], it is necessary to go thought a several steps process, the main points of which being:

1. identifying a unpredictable physical phenomenon that can be exploited in the operation of the electronic device and have a statistical model for this phenomenon;

2. design a TRNG using this phenomenon and have a stochastic model of the TRNG depending on parameters and giving the distribution of the output bits of the TRNG;

3. measure the parameters of the random phenomenons and using the stochastic model of the TRNG compute its entropy rate.

As far as we know, there is not much choice for a technological platform for which a simple and reliable implementation of all the steps of the preceding process in known. One known possibility uses the instability of the propagation time of the electric signal across logic gates called time jitter. It can be amplified in a so called ring oscillators which is composed of a sequence of inverters and delay elements connected in ring. A ring oscillator produces, when enabled, a clock signal the phase of which randomly fluctuate causing the so called phase jitter. The phase jitter can be exploited to produce random numbers. An example of simple well known design to do so, called Elementary Ring Oscillator based TRNG (ERO-TRNG) [4] is composed of two ring oscillators O_1 and O_2, the signal of O_2 sampling that of O_1 with a type-D flipflop. More complex designs such as MRO-TRNG are described in [4].

The phase jitter is the result of several physical phenomenons or noises with different statistical properties. The paper [3] describes a statistical model to compute contribution of the thermal noise to the entropy rate of the TRNG. Since the thermal noise is statistically independent of the other sources of noises, this model, based on a unidimensional Weiner process, is sufficient to compute a lower bound on the entropy rate of a oscillator based TRNG. The statistical parameters of this model are given by the drift λ and volatility σ of a Weiner process [3]. Measuring the physical quantity related to these parameters is one of the most challenging part of the design of an oscillator based TRNG with a good level of certainty on its security. One difficulty lies in the fact that it is necessary to distinguish the different components of the phase jitter (see [3] for a more in depth discussion on this subject). A way to distinguish the thermal noise component of the phase jitter is to measure it at high frequency where it is predominant over other noise sources.

The phase jitter can be measured externally, the clock signal being analyzed by an oscilloscope or with the internal method described in [2], the measure on the analog signal is done inside the chip and only numerical data are processed after this measure. The problem with the external method is that the acquisition chain add a lot of perturbation on the analog signals which make the measure less precise. It is also not practical because one can not use an oscilloscope on the assembly line of the chip. The second method is in theory far more accurate and practical. Still, in most cases, in practise, it does not

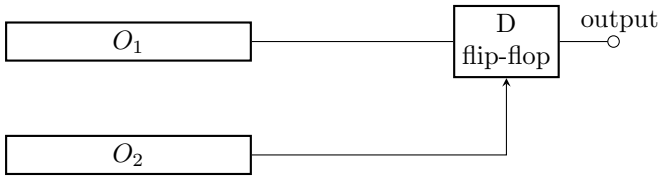

Figure 1: *Scheme of an elementary OJMD*

achieve the goals displayed in [2] to have a reliable and precise measure, simple enough to be embedded inside a chip (including the computational steps necessary to recover the parameters λ and σ). Actually, the good operation of the method depends on the choice of parameters and good configuration of the oscillators which is non trivial, not always possible and not explained. With a random choice of parameters and configurations in most cases the internal measure will lead to poor estimation of the statistical parameters of the thermal noise component of the phase jitter. This can lead to an overestimation of the entropy rate of the TRNG and thus a security vulnerability. The aim of this paper is to point to and explain the shortcomings of the method [2] and explain how to tweak it to overcome these problems.

The paper in organized as follows. In Section 2, we briefly recall the method of [2] in order to set the notations for the rest of the paper and explain in Section 3 the shortcomings of the method. In Section 4, we introduction improvements based an new algorithm to overcome the problems previously described. In Section 5, we present the outcome of our hardware implementation and extensive experiment of the new method to show its reliability. Then, we build upon the previous results to explain how to embed the internal measure method.

2 The internal measure method : summary and shortcomings

In this section, we briefly recall the method of [2] in order to fix the notations, explain its shortcomings and hint at our improvements. In [2], the authors show that what they call an elementary oscillator-based TRNG can be seen at the same time, for a different mode of operation, as an efficient measurement device for the phase jitter.

In order to simplify the presentation, we only describe the elementary oscillator-based TRNG in its configuration for measuring the phase jitter and call it an oscillator-based jitter measurement device (OJMD). An OJMD is composed of two oscillators, O_i for $i = 1, 2$. The output clock signal of O_2 is used to determine the sampling times of O_1. The sampling unit can be a synchronous D flip-flop (see Figure 1). In this case, the sample times correspond to the rising edges of the signal of O_2.

For $i = 1, 2$, the output signal of O_i is given by a quasi-periodic function $s_i(t)$ of time t that takes the form

$$s_i(t) = f_{\alpha_i}(\omega_i(t + \xi_i(t))), \tag{1}$$

where for $\alpha \in [0,1)$, we define f_α as the real valued 1-periodic function such that $f_\alpha(x) = 1$ for all $0 < x < \alpha$ and $f_\alpha(x) = 0$ for $\alpha < x < 1$, and $f_\alpha(0) = f_\alpha(\alpha) = 1/2$. The parameter α_i accounts for the duty cycle of $s_i(t)$. Up to replacing f_{α_i} by $1 - f_{\alpha_i}$, we will always suppose in the following that $\alpha_i \in [0, 0.5)$. For $i = 1, 2$, ω_i is the *mean frequency* of the signal $s_i(t)$, $\phi_i(t) = \omega_i(t + \xi_i(t))$ is its *phase* and the function $\xi_i(t)$ represents its *absolute phase drift* caused by the time jitter that we call phase jitter. Similarly, $T_i = 1/\omega_i$ is the mean period of $s_i(t)$.

Following [3], for $i = 1, 2$, we model the evolution of the total phase $\phi_i(t) = \omega_i(t+\xi_i(t))$ from Eq. (1), i.e. the phase of a ring oscillator O_i subject to the thermal noise, using a stationary Wiener stochastic process $\Phi_i(t)$ with drift $\mu_i > 0$ and volatility $\sigma_i^2 > 0$. In other words, for any time $t \geq t_0$, the phase $\Phi_i(t)$ conditioned by the value $\Phi_i(t_0) = \phi(t_0)$ follows a Gaussian distribution of expected value $\phi_i(t_0)+\mu_i(t-t_0)$ and variance $\sigma_i^2(t-t_0)$.

In [3, Appendix C], it is shown that, under general hypothesis always fulfilled in practise, the output of oscillator O_1 with duty cycle α_1 drift T_1 and volatility σ_1^2, sampled at time intervals determined by oscillator O_2 with drift T_2 and volatility σ_2^2 produces the same distribution of output bits as that of a stable clock signal (a jitter-free signal) with period duty cycle 0.5 and period T_2/T_1 sampling an oscillator O_1' with duty cycle α_1, drift 0 and volatility σ where $\sigma = \left(\frac{T_2}{T_1}\right)^2 \sigma_2^2 + \sigma_1^2$.

We call the triple:

$$(\alpha = \alpha_1, \mu = T_2/T_1 \mod 1, \sigma^2 = \left(\frac{T_2}{T_1}\right)^2 \sigma_2^2 + \sigma_1^2) \tag{2}$$

the statistical parameters of the OJMD composed of the oscillators O_1 and O_2. The knowledge of these parameters are enough to reproduce the distribution of the output bits $b = (b_j)_{j=1,\ldots,n}$ at sampling times $(t_j)_{j=1,\ldots,n}$ of the OJMD. Actually, the internal state of a OJMD with parameters (α, μ, σ^2) can be represented by a phase $\phi(t)$ depending on t. The output at time t of the OJMD knowing the phase according to (1) is $f_\alpha(\phi(t))$. As said before, the evolution of $\phi(t)$ is given by a Wiener process $\Phi(t)$ over the interval $[0, 1]$ with drift μ and volatility σ^2. So, from the knowledge of $\phi(t_0)$ at time t_0, at time $t > t_0$, the phase at t is drawn at random following the Gaussian distribution with expected value μ and standard deviation σ. The Algorithm 1 uses this in order to produce a series of bits following the distribution of a OJMD with given parameters (α, μ, σ^2).

In [2], the authors describe a method, which we briefly recall, to measure the statistical parameters (α, μ, σ^2) of a OJMD from the knowledge of a sample of its output bits. For $n \in \mathbb{N}^*$, let $b = (b_j)_{j\in\{1,\ldots,n\}}$ be the output bits sequence corresponding to the sampling of O_1 at times $(t_j = t_0 + j\mu)_{j\in\{1,\ldots,n\}}$ given by the rising edges of O_2 as depicted in Fig. 1. For any $S \subset \{1, \ldots, n\}$, we define

$$\mathbb{P}_S\{b_j \neq b_{j+M}\} = \frac{\#\{j \in S | b_j \neq b_{j+M}\}}{\#S}.$$

The method of [2] is based on three facts that we state without recalling all the conditions upon which they rely for the sake of simplicity and because in the following we are going to discuss these conditions. Let $N \in \mathbb{N}$ and for $i \in \{1, \ldots, n - N + 1\}$, we set $S_i = \{i, \ldots, i + N - 1\}$.

Algorithm 1: Algorithm to simulate a OJMD subject only to the thermal noise

input :

- (α, μ, σ^2) the statistical parameters of the OJMD;
- n a number of output bits;
- $(t_j)_{j=1,\dots,n}$ sampling times.

output: $(b_j)_{j=1,\dots,n}$ output bits.

$\phi = 0.5$ `/* Comment Initial relative phase */`
$t_0 \leftarrow t_1 - 1$;
for $i = 1$ *to* n **do**
  Draw a random x following the distribution $\tilde{G}(\mu(t_i - t_{i-1}), \sqrt{\sigma^2(t_i - t_{i-1})})$;
  $\phi \leftarrow (\phi + x) \mod 1$;
  if $\phi < \alpha$ **then**
    $b_i \leftarrow 1$;
  else
    $b_i \leftarrow 0$;
  end
end
return $(b_i)_{i=1,\dots,n}$.

Fact 1. *[2, Fact 1] Under certain hypothesis the set:*

$$S_{N,M} = \{c_i = \frac{1}{2}\mathbb{P}_{S_i}\{b_j \neq b_{j+M}\}\}_{i\in\{1,\dots,n-N-M+1\}}$$

is a sample drawn following the probability density function $D(M)(x)$ such that for all $a, b \in \mathbb{R}$, $\int_a^b D(M)(x)dx = \mathbb{P}\{(\Phi(t_0 + M\mu) \leq x | \Phi(t_0)) = x_0 \in [a, b]\}$ where $\Phi(t)$ is the Wieiner process governing the evolution of the phase $\phi(t)$ of the OJMD with statistical parameters (α, μ, σ^2).

Using Fact 1, one can recover σ^2 by computing the empirical variance $V(S_{N,M})$ of $S_{N,M}$ with any variance estimator such as:

$$V(S_{N,M}) = \frac{1}{K}\sum_{i=1}^{K} c_i^2 - (\frac{1}{K}\sum_{i=1}^{K} c_i^2).$$

We remark that, if we make the hypothesis that O_1 and O_2 are only subject to the thermal noise component of the phase jitter, $V(S_{N,1}) = \sigma^2$ and that $V(S_{N,M}) = MV(S_{N,1})$, so that in order to improve the measure of σ^2, one can compute a sample of $V(S_{N,M})$ for different values of M and recover $\sigma^2 = V(S_{N,1})$ as the slope of the line obtained as the linear regression of the samples. If M is small enough, the thermal noise will be preponderant over other noise sources so that the behaviour of $V(S_{N,M})$ as a function of M will be linear that the slope of $V(S_{N,M})$ gives σ^2.

Fact 2. *[2, Fact 2] Under certain hypothesis, we have for any $S \subset S$*

$$\frac{1}{2}\mathbb{P}_S\{b_i \neq b_{i+1}\} = \begin{cases} T_2/T_1 \mod 1 \text{ if } T_2/T_1 < 0.5 \\ 1 - T_2/T_1 \mod 1 \text{ otherwise.} \end{cases} \tag{3}$$

Fact 2 immediately gives a way to recover $\mu = T_2/T_1 \mod 1$.

Fact 3. *[2, Fact 1] Under certain hypothesis, we can obtain the duty cicle of the oscillator O_1 of the OJMD as:*

$$\alpha = \frac{\sum_{i=1}^{n} b_i}{n}. \tag{4}$$

It is clear that together Fact 1, Fact 2 and Fact 3, in principle, allow to recover the statistical parameters (α, μ, σ^2) of the OJMD. Nonetheless, as stated, these facts depend on hypothesis that are not always easy to fullfil and control. We are going to see in the next section that if these conditions are not verified, the precision of the measures may be greatly affected resulting in some cases in absurd outcomes.

3 Shortcomings of the method

We first treat the case of Fact 3, which is the easiest. Recall that $(t_j)_{j=1,\dots,n}$ are the sampling times corresponding to the rising edges of the signal produced by O_2. Then Fact 3 is true as long as $(t_i \mod T_1)$ follows a uniform distribution in $[0, T_1]$. This property is guarantied over a long period of time by the jitter phenomenon even if we suppose it to be very small. By the central limit theorem, the precision of the measure is in the order of $10^{\log_{10}(n)/2}$ with high probability.

As for Fact 2, it is true as long as $(t_i \mod T_1)$ follows a uniform distribution in $[0, T_1]$. Again, the precision of the measure is, with high probability, in the order of $10^{\log_{10}(n)/2}$.

The analysis of Fact 1 is much more delicate than the other facts in part because it depends on the choices of the parameters N and M which may affect its outcome. In order to assess its precision and the conditions upon which it works, we have made simulations using Algorithm 1 to produce series of bits $(b_j)_{j\in\{1,\dots,n\}}$ by a simulated OJMD with well chosen statistical parameters (α, μ, σ^2) for the thermal noise and used Fact 1 to recover σ^2 with different choices of N, M.

A first observation is that the line obtained as the linear regression of the samples $V(S_{N,M})$ does not pass through the origin as it should in theory: its equation is of the form $\sigma^2 x + b = 0$ where in general b in non-zero. Using simulation it is easily seen that this is due to the quantization error when approximating $\phi(t_0 + (M+i)\mu) - \phi(t_0 + i\mu)$ by c_i. This means in particular that one has to compute $V(S_{N,M})$ for at least two values of M to be able to recover σ^2.

In a first experiment, we have used Algorithm 1 with parameters $(\alpha = 0.5, \mu = 0.3376, \sigma^2 = 10^{-6})$. We have computed the samples $(S_{117,M})$ for $M = (300 + 5j)_{j=0,\dots,49}$ and plotted in Figure 2 the variance of these samples together with the line L_{σ^2} with slope $10^{-6} = \sigma^2$ that we should recover. We see that although most points $V(S_{117,M})$ fit correctly with L_{σ^2}, there are a lot of outliers. For instance, although $V(S_{117,540})$ is very close to L_{σ^2}, $V(S_{117,545})$ is widely apart.

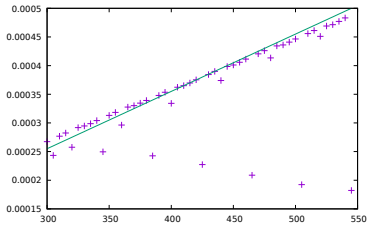

Figure 2: $V(S_{117,M})$ *for* $M = (300 + 5j)_{j=0,\ldots,49}$

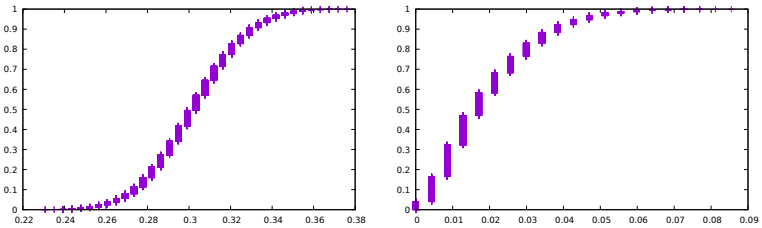

Figure 3: *Distribution of values* $V(S_{117,540})$ *(left) and* $V(S_{117,545})$ *(right)*

If we look in Figure 3 at the cumulative distribution function of $S_{117,540}$, we see that it is well approximating the cumulative distribution function of a Gaussian distribution as it should be for a Weiner process. This is not the case of the cumulative distribution function of $S_{117,545}$ on the right hand of Figure 3 where we note that the left tail of the expected Gaussian distribution is flattened.

How can we explain that ? In fact, if we have a closer look at the statement of [2, Fact 1], we see that the c_i are not drawn following the probability density function $D(M)(x)$ but rather that of $\psi_\alpha(D(M))$ where ψ_α is the function defined as:

$$\begin{aligned} \psi_\alpha(x) &= 2x \quad \text{if} \quad 0 \leq x \leq \alpha, \\ \psi_\alpha(x) &= 2\alpha \quad \text{if} \quad \alpha \leq x \leq 1-\alpha, \\ \psi_\alpha(x) &= 2-2x \quad \text{if} \quad 1-\alpha \leq x \leq 1. \end{aligned} \tag{5}$$

Using this we can explain the shape of the distributions obtained in Figure 3. In the left hand figure, the Gaussian distribution is contained in the interval $[0, \alpha]$ so that it is just stretched by a factor 2 by the function ϕ_α. In the right hand figure, the expected value of the Gaussian distribution is in 0 so that it is folded by ψ_α. This explains why, in the second case, the computation of the variance of the distribution is far lower than the expected one thus the corresponding abnormal point in the Figure 2. This problem occurs whenever the ϵ-support of the probability density function $D(M)$ intersects the set $\{0, \alpha\}$. By ϵ-support, we mean the smallest interval I centered around the expected value of $D(M)$ such that $\int_I D(M) \geq 1 - \epsilon$ for a small $\epsilon > 0$. But the mean value of $D(M)$ in $[0, 1]$ is given by $MT_2 \mod T_1$ so that it is fixed by the parameter M. Thus we see that whether the empirical distribution of $S_{N,M}$ is a good approximation of $D(M)$ depends of a choice of M that is not explained in [2].

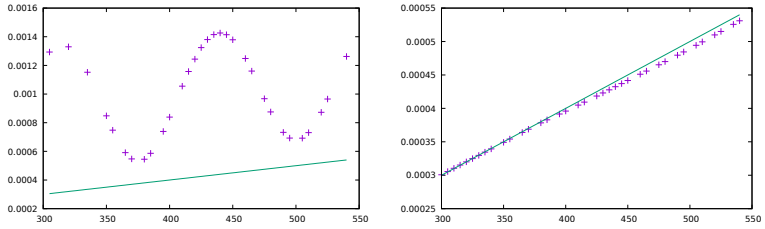

Figure 4: *Example of distribution of values* $V(S_{80,M})$ *(left) and* $V(S_{15,M})$ *(right) for* $M = (300 + 5j)_{j=0,\ldots,49}$

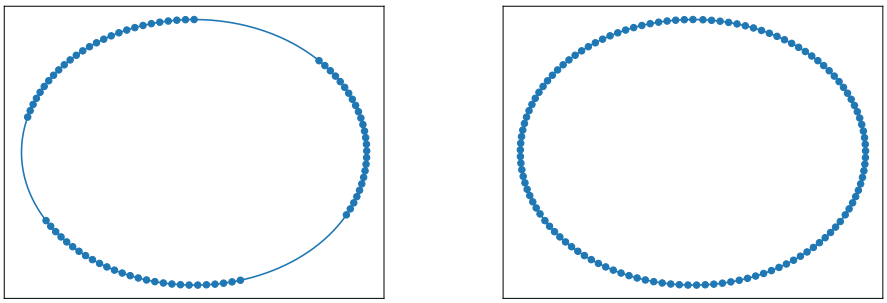

Figure 5: *Distribution of* $\{i\mu \mod 1, i = 1, \ldots N\}$*, for* $N = 80$ *(left) and* $N = 125$ *(right)*

We made a second experiment to illustrate the sensitivity of the method of [2] with the N parameter of Fact 1. We have generated a bit sequence using Algorithm 1 with parameters $(\alpha = 0.5, \mu = 0.332, \sigma^2 = 10^{-6})$. We have computed the samples $(S_{80,M})$ and $(S_{125,M})$ for $M = (300 + 5j)_{j=0,\dots,49}$ and plotted in Figure 5 the variance of these samples together with the line L_{σ^2} with slope $10^{-6} = \sigma^2$ that we should recover.

We see that $V(S_{125,M})$ fits correctly with the expected L_{σ^2} but that $V(S_{80,M})$ does not even looks like an affine law. The problem here is that Fact 1, rely on the hypothesis that the distribution of $\{i\mu \mod 1, i = 1, \dots, N\}$ is ϵ-uniform for a small positive ϵ. This means that the number of samples of $\{i\mu \mod 1, i = 1, \dots, N\}$ in the interval $[a, b]$ over N is close (up to a small ϵ) to the expected one if the distribution of the $\{\xi\mu \mod 1, i = 1, \dots, N\}$ was uniform in the interval $[0, 1]$ that is the size of the interval $[a, b]$.

We see in Figure 4 that the distribution of $\{i\mu \mod 1, i = 1, \dots, N\}$ is uniform in $[0, 1]$ (represented by a circle in the figure) if $N = 80$ but this is not the case if $N = 40$. This time we see that whether the empirical distribution of $S_{N,M}$ is a good approximation of $D(M)$ also depends in a crucial manner on the choice of N.

4 Improvements

In the previous section, we have described cases when the distribution of $S_{N,M}$ that we can compute is not a good approximation of that of $D(M)$ that we want to recover. A first problem affects the choice of M and the second that of N so that if one does not take care when choosing the parameters of the method of [2], the outcome may be very far from the result that we look for. In this section, we explain how to tweak the jitter measurement method of [2] to overcome these defects and so to improve its precision and reliability.

4.1 Outlier detection and mitigation

We have seen that the distribution of the $S_{N,M}$ that we compute does not follow the probability density function $D(M)$ but rather that of $\psi_\alpha(D(M))$ where ψ_α is defined by (5).

In the case that the ϵ-support of the distribution $D(M)$ intersects $\{0, \alpha\}$ the distribution of the $S_{N,M}$ will not be, in general, a good approximation of $D(M)$ up to a linear factor. In order to avoid this, one can filter out bad values of M so that the ϵ-support of $S_{N,M}$ does not meet $\{0, \alpha\}$. We suppose that σ is very small compared to 1 which is the case in most of the measures that we have done on a wide range of different technologies.

A first idea to filter out bad values of M is to choose M so that

$$\sqrt{\sigma^2 M} < \epsilon_0 \tag{6}$$

$$|\mathbb{E}(D(M)) - \alpha/2| < \epsilon_0, \tag{7}$$

for $\epsilon_0 > 0$ much smaller than α (for instance take $\epsilon_0 = \alpha/4$). Actually, if the variance of $D(M)$ is small and its expected value is far from $\{0, \alpha\}$ then the probability that the ϵ-support of $D(M)$ gets across $\{0, \alpha\}$ is small. We remark that, as one can evaluate $T_2/T_1 \mod 1$ and α using respectively Facts 2 and 3, it is possible to recover $|\mathbb{E}(D(M)) - \alpha/2|$ as

$[MT_2/T_1 \mod 1-\alpha/2|$ and so choosing a M verifying the second condition. Nonetheless, as don't know σ^2 because this is exactly what we want to measure, it is difficult to verify that the first condition is fulfilled. So the method works only if one knows an order of magnitude of σ^2 and take margins. The problem is then that we are going to discard a lot of values of M suitable for the computation of σ^2. By being to much selective on M we take the risk to eliminate all possible values or diminish drastically the number of points used to compute a linear regression and so lose precision. We would like to be able to asses if the distribution $S_{N,M}$ actually gets across $\{0, \alpha\}$ without an a priori knowledge of σ^2.

In order to explain how to do it, we suppose to simplify the explanation that $S_{N,M}$ gets across 0 (the case of α is similar). Let $c_i = \frac{1}{2}\mathbb{P}_{S_i}\{b_j \neq b_{j+M}\}$ for $i \in \{1, \ldots, n-N-M+1\}$ and we recall that $S_i = \{i, \ldots, i+N-1\}$. Let $\delta(M,t) = \phi(t+M\mu) - \phi(t)$. As the phase jitter per period is small, we have:

$$|\delta(M, t_0 + i\mu) - \delta(M, t_0 + (i-1)\mu)| < \epsilon_1, \tag{8}$$

for a small ϵ_1 (of the order of $\sqrt{\sigma^2 T_2}$). Moreover, if the conditions of Fact 1 are fullfiled, there exists a small $\epsilon_2(M) > 0$ such that for all $i \in \{1, \ldots, n-N-M\}$,

$$|x - \delta(M, t_0 + i\mu)| < \epsilon_2(M), \tag{9}$$

for $x \in \{c_i, 1-c_i\}$. When c_i is far from 0, c_i and $1-c_i$ are far apart from each other so that it is easy to decide which one is the correct approximation of $\delta(M, t_0 + i\mu)$ by induction on i: for c_0 we can choose at random any of c_0 and $1-c_0$, this will act by -1 on the computed distribution of the $\delta(M, t_0 + i\mu)$ for $i = 1, \ldots, N$ without affecting its variance. Then suppose that we have chosen a good approximation $y \in \{c_{i-1}, 1-c_{i-1}\}$ of $\delta(M, t_0 + (i-1)\mu)$ then by equations (8) and (9), $x \in \{c_i, 1-c_i\}$ is a good approximation of $\delta(M, t_0 + i\mu)$ if it verities $|x-y| < \epsilon_1 + 2\epsilon_2(M)$.

But when c_i is near to 0, we don't have a criterion to decide which of c_i or $1-c_i$ is the good approximation of $\delta(M, t_0 + i\mu)$ so we can not use the value of c_i to detect if $\delta(M, t_0 + i\mu)$ crosses 0 for different values of i. In order to improve that, using the fact that, in practise, the jitter by period $\sqrt{\sigma^2 T_2}$ is small, for $\lambda = \pm 1$, we have for all t_0: $|\phi(t_0 + M\mu) - (\phi(t_0 + (M-\lambda)\mu) + \lambda\mu)| < \epsilon_1$ for a small ϵ_1 of the order of $\sqrt{\sigma^2 T}$. This means that:

$$|\delta(M, t_0 + i\mu) - (\delta(M - \lambda, t_0 + i\mu) + \lambda\mu)| < \epsilon_1. \tag{10}$$

Based on this remark, keeping the definition for $\epsilon_1, \epsilon_2(M)$ from above, we have the following fact:

Fact 4. *Let $\lambda = \pm 1$ be an integer, let $c_i = \frac{1}{2}\mathbb{P}_{S_i}\{b_j \neq b_{j+M}\}$ and $c'_i = \frac{1}{2}\mathbb{P}_{S_i}\{b_j \neq b_{j+M-\lambda}\}$. Set $\epsilon = \epsilon_1 + 2max(\epsilon_2(M), \epsilon_2(M \pm \lambda))$. If we have*

$$|\lambda\mu \mod 1| > \epsilon, |\lambda\mu - \frac{1-2c_i}{2} \mod 1| > \epsilon, \tag{11}$$

there is a unique pair (x, y) where $x \in \{c_i, 1-c_i\}$ and $y \in \{c'_i + \lambda\mu, 1-c'_i + \lambda\mu\}$ such that $|x-y| < \epsilon$.

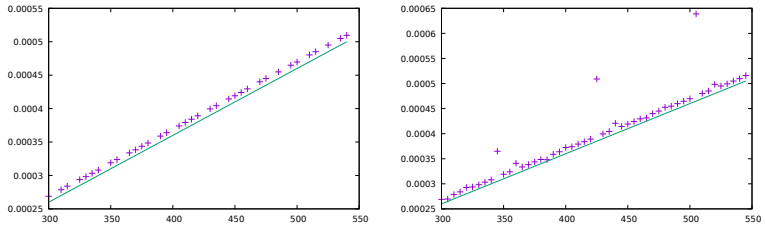

Figure 6: $V(S_{117,M})$ *for* $M = (300 + 5j)_{j=0,\dots,49}$ *discarding outliers (left) and by gluing distributions (right)*

Proof. The existence of the pair (x, y) is an immediate consequence of the triangular inequality and the fact that for a $x \in \{c'_i, 1 - c'_i\}$, $|x - \delta(M - \lambda, t_0 + i\mu)| < \epsilon_2$, for a $y \in \{c_i, 1-c_i\}$, $|y-\delta(M, t_0+i\mu)| < \epsilon_2$ and moreover that $|\delta(M, t_0+i\mu)-\delta(M-\lambda, t_0+i\mu)+\lambda\eta \mod 1| < \epsilon_1$.

For the unicity, suppose that $|c'_i - c_i + \lambda\mu| < \epsilon$. If $|1 - c'_i - c_i + \lambda\mu \mod 1| < \epsilon$ then $|\lambda\mu - \frac{1-2c_i}{2} \mod 1| < \epsilon$. If $|1 - c'_i - 1 - c_i + \lambda\mu| < \epsilon$, then $|\lambda\mu - 0 \mod 1| < \epsilon$. The case, $|c'_i - 1 - c_i + \lambda\mu| < \epsilon$ is similar. □

In order to use the preceding Fact, we remark that we can obtain a good approximation of μ using Fact 2 and then use it to check if the conditions (11) are fulfilled (in practise, one can take safely $\epsilon = 1/10$). Then Fact 4 allows to lift the ambiguity about which of c_i and $1 - c_i$ is a good approximation of $\delta(M, t_0 + i\mu)$ except in the rare cases when conditions (11) are not fulfilled. In this case, one can redo the computations with $\lambda = -1$. This gives a way to detect when c_i gets across 0.

We can use it either to detect when a c_i is likely to be an outlier and discard it as in Algorithm 3. It is also possible to compute c_i taking into account of the fact that it has crossed 0 or α as in Algorithm 4. Note that in this last algorithm the distribution of $c[i]$ that we compute can get across 0 or α but the $\mathbb{P}_S(b_j \neq b_{j+M})$ is always in $[0, 1]$ so that we have to glue them together using the integer offset that is updated by plus or minus 1 when the $c[i]$ goes through 0 or 1.

We have tested the two algorithms with the Wiener process $W(\alpha = 0.5, \mu = 0.3376, \sigma = 10^{-6})$ and we have computed the samples $(S_{117,M})$ for $M = (300 + 5j)_{j=0,\dots 49}$ exactly as for Figure 2. We remark in Figure 6 (left) that the new method allows to detect and remove all the outlier. When we use it to glue the distribution of c_i together we still have some outlier but far less than with the old method as it is shown in Figure 6 (right). These outlier are due to the fact that it may happen very rarely that $\delta(M, t_0 + i\mu)$ and $\delta(M - \lambda, t_0 + i\mu)$ are far enough so that we are wrong when using it to decide whether to choose c_i or $1 - c_i$.

4.2 Continued fractions and ϵ-uniformity

In this section, we consider that the interval $[0, 1[$ is a circle by identifying 0 and 1. We say that $T \subset [0, 1[$ is an interval if $\cup_{t\in\mathbb{Z}}(T + t)$ is an interval of $\mathbb{R}$.

Algorithm 2: *Algorithm to compute the probability density function $D(M)$*

input :

- The output sequence $[b_1, \ldots, b_n]$ of an OJMD;
- $\eta = T_1/T_2 \mod 1$, λ a small integer;
- M and N integers.

output: Fail or $c[i]$ which have the same distribution as $D(M)$.

$i \leftarrow 0$;
for i *in range*$(n - M - N)$ **do**
  $S_i \leftarrow [i + 1, \ldots, i + N]$;
  $c = \mathbb{P}_{S_i}(b_j \neq b_{j+M})$;
  $c' = \mathbb{P}_{S_i}(b_j \neq b_{j+M-\lambda}) + \eta\lambda$;
  $C = [c, 1 - c]$;
  $C' = [c', 1 - c']$;
  Let (k, l) such that $|C[k] - C'[l]| = \min\{|x - y|, x \in C, y \in C'\}$;
  $c[i] = C[k]$;
end
return $c[i]$;

We have seen in Section 2 that the outcome of Fact 1 can be affected by the choice of N. In order to have a better understanding of the relation between N and the precision of the computation of σ^2 by Fact 1, we recall some definitions and statements from [3].

For $x, y \in [0, 1[$, let $d(x, y) = \min(|x - y|, 1 - |x - y|)$. If I is an interval of $[0, 1[$, we define the diameter of I denoted by $d(I)$ the quantity $\max(d(x, y), x, y \in I)$. Let K be a finite subset of the interval $[0, 1[$, we say that K is ϵ-uniform if for all $[a, b] \subset [0, 1]$, we have:

$$|\frac{\sharp K \cap [a, b]}{\sharp K} - (b - a)| \leq \epsilon. \tag{12}$$

In the following, for any K finite subset of $[0, 1[$, we denote by $\epsilon(K)$ the minimum of the set $\mathcal{L} = \{\epsilon > 0 | K \text{ is } \epsilon\text{-uniform}\}$ (such a minimum exists since $\mathcal{L}$ is closed and bounded from below by 0). For any i positive integer, we let $\kappa(i) = \phi(t_0 + (i + M)\mu) - \phi(t_0 + i\mu) \mod 1 \in [0, 1[$. We set $S_i = \{i, \ldots, i + N - 1\}$ for $i = 1, \ldots, n - N + 1$ and for $i = 1, \ldots, n - N - M + 1$, let $K_i = \{\kappa(j), j \in S_i\}$, $\mathbb{E}_{S_i}(\kappa) = \frac{1}{N}\sum_{j \in S_i} \kappa(j)$. If we have a closer look at the statement of [3, Fact 1], we see that:

$$|\frac{1}{2}\mathbb{P}_{S_i}\{b_j \neq b_{j+M}\} - min(\mathbb{E}_{S_i}(\kappa), 1 - \mathbb{E}_{S_i}(\kappa)| < \epsilon(K_i).$$

In other words, the points that we compute in order to recover the distribution $D(M)$ are approximation up to $\epsilon(K_i)$ of the real points $\mathbb{E}_{S_i}(\kappa(i))$ or $1 - \mathbb{E}_{S_i}(\kappa(i))$. We would like to choose N in order to obtain the smallest possible $\epsilon(K_i)$. A first question is: when K run through the set of finite subsets of cardinality N of $[0, 1[$, what is the span of $\epsilon(K)$ and for what configuration of K $\epsilon(K)$ is the smallest possible. This question is answered

Algorithm 3: *Algorithm to compute the probability density function $D(M)$*

input :

- The output sequence $[b_1, \ldots, b_n]$ of an OJMD;
- $\eta = T_1/T_2 \mod 1$, λ a small integer;
- K, M and N integers.

output: $c[i]$ which have the same distribution as $D(M)$.

for $i = 0, \ldots, K$ **do**
  offset $\leftarrow 0$;
  $S_i \leftarrow [Ni + 1, \ldots, Ni + N]$;
  $c = \mathbb{P}_S(b_j \neq b_{j+M})$;
  $c' = \mathbb{P}_S(b_j \neq b_{j+M-\lambda}) + \eta\lambda$;
  $\nu \leftarrow \{\min(x, y)\, x \in \{c, 1-c\}, y \in \{c', 1-c'\}\}$;
  if $\nu - c[i-1] > 0.5$ **then**
    offset $\leftarrow$ offset $+ 1$;
  end
  if $\nu - c[i-1] < -0.5$ **then**
    offset $\leftarrow$ offset $- 1$;
  end
  $c[i] = c +$ offset;
end
return $c[i]$;

by the following lemma which explains that $\epsilon(K)$ is optimal when the points of K are uniformly distributed in $[0,1[$. We state also the basic invariance of $\epsilon(K)$ with respect to translations.

Lemma 1. *Let $K(N)$ by a finite subset of $[0,1[$ of cardinality N, we have:*

- *$\epsilon(K(N)) = \epsilon(K(N)+t)$ for any $t \in \mathbb{R}$ (invariance of $\epsilon(K(N))$ by translation by t), $\epsilon(K(N)) < 1$;*
- *if there exists $\zeta > 0$ and I an interval of $[0,1[$ such that $d(I) = \zeta$, $I \cap K(N) = \emptyset$ then $\epsilon(K(N)) \geq \zeta$;*
- *$\epsilon(\{i/N \mod 1, i = 0, \ldots, N-1\}) = 1/N$.*

If $K(N)$ has cardinality N then $\epsilon(K(N)) \in [1/N, 1[$ and because of the preceding these bounds are optimal.

Proof. The first two claims of the Lemma are clear.

Let $K = \{i/N \mod 1, i \in \{0, \ldots, N-1\}$. By applying the second claim to $]0, 1/N[\cap K = \emptyset$, we have $\epsilon(K) \geq 1/N$. Let $I \subset [0,1[$ be an interval, we have

$$\lfloor Nd(I) \rfloor / N \leq \sharp(K \cap I)/N \leq \lfloor (Nd(I) \rfloor + 1)/N \tag{13}$$

and moreover,

$$\lfloor Nd(I) \rfloor / N \leq d(I) \leq \lfloor Nd(I) \rfloor / N + 1/N. \tag{14}$$

Using Equations (13) and (14), we obtain that $\epsilon(K) \leq 1/N$.

If K has cardinality N then $\min(\{d(x,y), x, y \in K\} \geq 1/N$ so that $\epsilon(K) \geq 1/N$ whence the last claim. □

As N is small, considering that the phase jitter is small during NT_2, we can suppose that $\phi(i\eta) = i\eta \mod 1$. So we have to study the ϵ-uniformity of the set $K(N) = \{i\eta \mod 1, i = 0, \ldots, N-1\}$. A basic remark is that if η is a rational, write it as P/N an irreducible fraction then $K(N) = \{i/N, i = 0, \ldots, N-1\}$ and we have seen in Lemma 1 that $\epsilon(K(N)) = 1/N$ and is minimal among all finite subsets of $[0,1[$ of cardinality N. However in our application, for typical values, η will be a real number known up to precision say 10^{-3} using for instance Fact 2 to compute η with a sample of 10^6 bits. So that η is of the from P/N with N of the order to 10^3 and we would like to choose N of the order to 10^2 so that N is small enough to comply with the hypothesis that the jitter during NT_2 is small. So we can not apply the preceding naive approach. But it seems reasonable to think that $\epsilon(K(N))$ will be the smallest when we choose N so that there exists a good rational approximation P/N of η. These good approximations are given by the theory of continued fractions (first introduced in [5] in the context of jitter measurement of oscillator based TRNG). We briefly recall the classical results and notations that we use in the following and refer the reader to [6] for a more in-depth introduction to continued fractions.

Let x be a real number, we denote by $[a_0, a_1, \ldots]$ where $a_0, a_1, \ldots$ are positive integers, the (possibly infinite) continued fraction representation of x that is:

$$x = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \ldots}}. \tag{15}$$

If $s = [a_0; a_1, \ldots]$ is a continued fraction and k a positive integer, we denote by $s_k = [a_0; a_1, \ldots, a_k]$ its k^{th} segment of by $\frac{s(p)_k}{s(q)_k}$ its k^{th} convergent. In other words, $\frac{s(p)_k}{s(q)_k}$ is the unique irreducible fraction representing $[a_0; a_1, \ldots, a_k]$. Let s_∞ be the real number represented by s. Denote by $d(s,k) = s(q)_k s_\infty - s(p)_k$ and more generally by $d(s,k,\lambda) = \lambda d(s,k+1) + d(s,k)$ for any $\lambda \in \{0, \ldots, a_{k+2}\}$ positive integer. We say, following [6], that a/b a fraction with $b > 0$ is a best approximation of second kind of a real number α if for all λ/μ fraction such that $\lambda/\mu \neq a/b$ and $0 < \mu < b$, we have $|b\alpha - a| < |\mu\alpha - \lambda|$.

We have gathered in the following proposition the main results of the theory of continued fraction that we use:

Proposition 1 ([6])**.** *We have:*

1. *[6][Theorem 4] for all k positive integer, $(-1)^k d(s,k) > 0$ and the sequence*
$$(-1)^k (d(s,k))_{k\in\mathbb{N}}$$
is decreasing;
2. *[6][Theorem 1]we have the relations $s(p)_k = a_k s(p)_{k-1} + s(p)_{k-2}$, $s(q)_k = a_k s(q)_{k-1} + s(q)_{k-2}$ for all $k \geq 2$;*
3. *[6][Theorem 16] an irreducible fraction λ/μ is a best approximation of second kind of s_∞ if and only if $\lambda/\mu = s(p)_k/s(q)_k$ for a certain k.*

It is clear that $d(s,k,0) = d(s,k)$ and by Proposition 1[2.], we have

$$d(s,k,a_{k+2}) = d(s,k+2). \tag{16}$$

We need a slight generalisation of the definition of approxiamtion of second kind:

Definition 1. *We say that a fraction a/b, $b > 0$ is a best positive (resp. negative) approximation of second kind of a real number α if for all λ/μ such that $\mu > 0$, $\lambda/\mu \neq a/b$, $0 < \mu < b$, we have $0 < b\alpha - a < \mu\alpha - \lambda$ (resp. $b\alpha - a > \mu\alpha - \lambda > 0$).*

With this definition, we can state the following lemma which will be useful:

Lemma 2. *Let $s = [a_0; a_1, \ldots]$ be the continued fraction representation of s_∞. We have:*

- *For all k, λ positive integer such that $\lambda \in \{0, \ldots, a_{k+2} - 1\}$, $(-1)^k d(s,k,\lambda) > 0$;*
- *The fraction λ/μ, $\mu > 0$ is a best positive (resp. negative) approxmation of second kind of s_∞ if and only if $\lambda = \lambda_0 s(p)_{k+1} + s(p)_k$ and $\mu = \lambda_0 s(q)_{k+1} + s(q)_k$ for $\lambda_0 \in \{0, \ldots, a_{k+2}\}$ and k even (resp. odd).*

Proof. For the first claim, it is true for $\lambda = 0$ by Proposition 1 since $d(s,k,0) = d(s,k)$ by definition. Next, we suppose in order to simplify the notations that k is even and leave the odd case to the reader. Suppose that for a $\lambda \in \{0, \ldots, a_{k+2} - 1\}$, we have $d(s,k,\lambda) < 0$. Let λ_0 be the smalest positive integer realizing this condition. Then $0 \in]d(s,k,\lambda_0 + 1), d(s,k,\lambda_0)[$ and we have $d(s,k,\lambda_0 + 1) - d(s,k,\lambda_0) = d(s,k+1) < 0$

by definition. This means that $|d(s,k,\lambda_0)| < |d(s,k+1)|$ and $s(q)_k + \lambda_0 s(q)_{k+1} < s(q)_{k+2}$. But this contradicts Proposition 1 saying the only the convergeants are best approximation of second kind.

We prove the second claim, by induction on k. We suppose it to be true for all $k \in \{0,\dots,k_0-1\}$, the case $k_0 = 0$ being trivial since $s(q)_0 = 1$. We do the induction in the case that k_0 is even the odd case being similar.

So we have to prove that for $\lambda_0 \in \{0,\dots,a_{k_0+2}-1\}$, $P(\lambda_0)/Q(\lambda_0)$ for $P(\lambda_0) = \lambda_0 s(p)_{k_0+1} + s(p)_{k_0}$ and $Q(\lambda_0) = \lambda_0 s(q)_{k_0+1} + s(q)_{k_0}$ is a best positive approximation of second kind and that any best positive approximation of second kind is of this form.

We do an induction on λ_0. For $\lambda_0 = 0$, the result is an immediate consequence of Proposition 1, because $Q(0)s_\infty - P(0) > 0$ and $P(0)/Q(0)$ is a best approximaiton of second kind. We make the inductive hypothesis that for $\lambda_0 < a_{k_0+2}-1$, $P(\lambda_0)/Q(\lambda_0)$ is a best positive approximation of second kind and we want to prove that $P(\lambda_0+1)/Q(\lambda_0+1)$ is the only next one. We suppose the contrary and let λ/μ, $\mu > 0$ be a rational such that

$$Q(\lambda_0)s_\infty - P(\lambda_0) > \mu s_\infty - \lambda > 0,\ Q(\lambda_0) < \mu < Q(\lambda_0+1). \tag{17}$$

Set $\lambda_1 = \lambda - P(\lambda_0)$ and $\mu_1 = \mu - Q(\lambda_0)$. Then by Equation (17) and using the fact that $d(s,k_0) > Q(\lambda_0)s_\infty - P(\lambda_0)$ (because $d(s,k_0+1) < 0$), we have: $0 > \mu_1 s_\infty - \lambda_1 > -s(d,k_0)$. Moreover, because of Equation (17), we have $\mu_1 < s(q)_{k_0+1}$ so that by Proposition 1, $0 > s(d,k_0+1) > \mu_1 s_\infty - \lambda_1$. We can gather all the provious in the inequalities:

$$0 > s(d,k_0+1) > \mu_1 s_\infty - \lambda_1 > -s(d,k_0). \tag{18}$$

But, by the induction hypothesis for $k_0 - 1$, we know that if λ_2/μ_2 is a best negative approximation of second kind of s_∞ with μ_2 the biggest such that $0 < \mu_2 < s(q)_{k_0+1}$, then $\mu_2 s_\infty - \lambda_2 = d(s,k_0-1,a_{k_0+1}-1) = d(s,k_0+1) - d(s,k_0)$ because of Equation (16). Comparing with Equation (18), we obtain:

$$0 > \mu_1 s_\infty - \lambda_1 > \mu_2 s_\infty - \lambda_2. \tag{19}$$

Recall that $\mu_1 < s(q)_{k_0+1}$, the Inequality (19) contractids the hypothesis that λ_2/μ_2 is a best negative approximation of second kind of s_∞ with μ_2 the biggest such that $\mu_2 < s(q)_{k_0+1}$.

We deduce that λ/μ verifying Equation (17) does not exists and so that $P(\lambda_0+1)/Q(\lambda_0+1)$ is the only next best positive approximation of second kind of s_∞ after $P(\lambda_0)/Q(\lambda_0)$. □

Let $\eta \in \mathbb{R}$ and $s = [a_0; a_1, \dots]$ be its continued fraction representation so that $s_\infty = \eta$. Let N be a positive integer and $i \in \{0, N-1\}$. We consider the finite set $K(N) = \{i\eta \mod 1, i = 0,\dots,N-1\} \subset [0,1[$. We suppose that the cardinal of $K(N)$ is N and we rank the elements of $K(N)$ by writing $K(N) = \{\kappa_j, j = 1,\dots,N-1\}$ such that $\kappa_j \le \kappa_{j+1}$ for $j = 0,\dots,N-2$. Let χ be the permutation of $\{0,\dots,N-1\}$ such that $\kappa_j = \chi(j)\eta \mod 1$. Let $s = [a_0; a_1, \dots]$ be the continued fraction representation of $s_\infty = \eta$. Denote by $\mathcal{D}^N$ the set of all $d(s,k,\lambda)$, for k a positive integer and $\lambda \in \{0,\dots,a_k-1\}$ such that $a_k s(q)_{k-1} + s(q)_{k-2} \le N$. It is clear from Proposition 1 that $\mathcal{D}^N$ is finite. In order to compute $\epsilon((K(N))$, we need to understand what are the quantities $d(\kappa_{j+1},\kappa_j)$. For this we denote by $\mathcal{I} = \{d(\kappa_{j+1},\kappa_j), j = 1,\dots,N-1\}$. We use the following Lemma :

Lemma 3. *For $j \in \{0, \dots N-1\}$, let $\mathcal{D}_j^N$ be the set of all $d(s,k,\lambda) \in \mathcal{D}^N$ such that*

$$\chi(j) + (-1)^k(\lambda s(q)_{k+1} + s(q)_k) \in \{0, \dots, N-1\}. \tag{20}$$

Let $d(s,k_0,\lambda_0)$ be the smallest element of $\mathcal{D}_j^N$. Then we have $\chi(j+1) = \chi(j) + (-1)^{k_0}(\lambda_0 s(q)_{k_0+1} + s(q)_{k_0})$ and $d(\kappa_{j+1}, \kappa_j) = (-1)^{k_0} d(s,k_0,\lambda_0)$.

Proof. Let $\mu = \chi(j+1) - \chi(j)$. We remark that $\mu > 0$ (resp. $\mu < 0$) if and only if there exists λ such that $\lambda/|\mu|$ is the best positive (resp. negative) approximation of second kind of η with $\mu < N - \chi(j)$ (resp. with $\mu + \chi(j+1) > 0$). Actually, if $\mu > 0$, let λ'/μ' be the best positive approximation of η with $\mu' < N - \chi(j)$. If $\mu' \neq \mu$ then $0 < \eta\mu' - \lambda' < \eta\mu - \lambda$. But this means that

$$\chi(j)\eta \mod 1 < (\chi(j) + \mu')\eta \mod 1 < \chi(j+1)\eta \mod 1 \tag{21}$$

with $\chi(j) + \mu' \in \{0, \dots, N-1\}$. But this is a contracdiction with the definition of χ. The negative case can be treated in the same manner.

□

Corollary 1. *We let $\mathcal{D}^N = \{d_i\}$ with $d_i < d_{i+1}$ and denote by $\mathcal{D}^N_{\leq i_0}$ the set $\{d_i, i \leq i_0\} \subset \mathcal{D}^N$. Let i_0 be the smallest index such that $\mathcal{D}^N_{\leq i_0}$ contains $d(s,k_1,\lambda_1)$, $d(s,k_2,\lambda_2)$ and such that:*

1. $(-1)^{k_1} + (-1)^{k_2} = 0$;
2. $|(\lambda_1 s(q)_{k_1+1} + s(q)_{k_1}) + (\lambda_2 s(q)_{k_2+1} + s(q)_{k_2})| \leq N$

Then we have $\mathcal{D}^N_{\leq i_0} \supset \mathcal{I}$.

Proof. For $j \in \{0, \dots, N-1\}$, we have to prove that $\chi(j+1) - \chi(j) = (-1)^k(\lambda s(q)_{k+1} + s(q)_k)$ for $d(s,k,\lambda) \in \mathcal{D}^N_{\leq i_0}$. But the two hypothesis of the Corollary ensure that there is $d(s,k,\lambda) \in \mathcal{D}^N_{\leq i_0}$ verifying (20) since $\chi(j) \in \{0, \dots, N-1\}$. Thus $\mathcal{D}^N_{\leq i_0}$ contains the smallest $d(s,k,\lambda)$ vérifying (20). By applying Lemma 3, we obtain the result. □

As the distance between two points of $K(N)$ are elements of $\mathcal{I} \subset \mathcal{D}^N_{\leq i_0}$ verifying the conditions of the previous Corollary, the set $K(N)$ is especially uniformly distributed in $[0,1[$ when there exists such a $\mathcal{D}^N_{\leq i_0}$ with small cardinality. The following Corollary tells that, it has the smallest possible cardinality which is 2 when N is a convergent not equal to s_∞ (this last case is covered by Lemma 1).

Corollary 2. *Suppose that $N = s(q)_k$ for $k \geq 2$ and $s_\infty \neq s(p)_k/s(q)_k$. Then, $\mathcal{D}^N_{\leq i_0} = \{d(s,k-2,\lambda), \lambda = 0, \dots, a_k - 1\}$ and $\mathcal{I} = \{|d(s,k-2)|, |d(s,k-2) + d(s,k)|\}$.*

Proof. We apply Corollary 1 with $d(s,k_1,\lambda_1) = d(s,k-1,0)$ and $d(s,k_2,\lambda_2) = d(s,k-2,a_k-1)$. Using Proposition 1, we have that $\sum_{i=1,2} \lambda_i s(q)_{k_i-1} + s(q)_{k_i-2} = s(q)_k = N$, by hypothesis and it is clear that $(-1)^{k_1} + (-1)^{k_2} = 0$. □

Under the hpothesis of the Corollary it is easy to see that $\epsilon(K(s(q)_k)) = d(s, k-2) + d(s, k)$.

Recall that η can be measured easily using the algorithm deduced from Fact 2. Then it is well known that the $s = [a_0; a_1, \ldots]$ can be obtained with a sequence of Euclidean divisions as it is explained in Algorithm 4. From the continued fraction representation of η it is trivial to obtain the set of convergent with the recurrence formulas of Proposition 1. In the same way, in our first experiment of Section 3, we have considered an OJMD with paremeters $(\alpha = 0.5, \mu = 0.3376, \sigma = 10^{-3})$. The convergents of 0.6752 are $[1, 1, 3, 37, 40, 117, 507]$ whence the choice of 117 in our computations.

Algorithm 4: *Algorithm to compute the continued fraction representation of ζ.*

input :

- $\zeta \in \mathbb{R}$;
- n a positive integer.

output: $s = [a_0; a_1, \ldots, a_M]$ the M^{th} remainder of the continued fraction representation of ζ.

```
x ← ζ;
s ← [];
for i = 0, ..., M do
    x_0 ← ⌊x⌋;
    r ← x − x_0;
    s ← s + [x_0];
    if r ≠ 0 then
        x=1/r;
    else
        Break;
    end
end
return s;
```

If we get back to the example presented in Section 2 for the oscillator with parameters $(\alpha = 0.5, \mu = 0.332, \sigma = 10^{-3})$. If we compute the successive $s(q)_k$ for the continued fraction such that $s_\infty = 0.664$, we obtain $[1, 1, 2, 3, 125, \ldots]$ which explains according to the previous Corollary why the distribution of the $\{i\mu \mod 1, i = 1, \ldots, N\}$ looks more uniformly distributed in Figure 4 for $N = 125$ than for $N = 80$ and thus the discrepancy in the result of the measure when we choose $N = 80$.

5 Embedded implementation of the measure

5.1 Implementation

We have implemented and tested the algorithms of this paper on a FPGA XILINX ARTIX-7 (28nm HPL technology of TSMC). We have implemented 32 rings oscillator O_i

for $i = 0, \ldots, 31$ of different length together with a fixed sampling ring oscillator O_{ech}. A cell is implemented on the FPGA by a LUT. If the number n of cells is pair (resp. odd), the ring oscillator is made of $n-1$ inverters and one AND (resp. NAND) cell to cycle the electric signal and enable the ring. We obtain 32 OJMD denoted OJMD_i for $i = 0, \ldots, 31$, where OJMD_i is made of the ring O_i of mean period T_i sampled by O_{ech} for mean period T_{ech} by the way of a type-D flip flop.

Constraints for positioning each element of the ring as well as bounds on time delay on the rings allow to control the reproducibility of the behaviour of the rings from an implementation to another. Note that we use a AND (or NAND) gate so that the ring is disabled when we set the configuration of the FPGA in order to avoid to have several clock edges per period. A double counter in implemented, one of them counting the edges of a clock the frequency of which is known and stabilized (for instance by a quartz) allows to measure the frequency of the rings.

An acquisition module is implemented on the FPGA. It allows to save 262 144 bits at the output of the each OJMD. These data are uploaded on a computer by the way of a UART link and saved in a file. We get 100 files for each OJMD. The experiments are done at ambient temperature and with the nominal voltage for the FPGA. For a complete characterisation of a oscillator based TRNG furthers experiments would be necessary to test the behavior each OJMD with respect to parameters which are known to have influence on its operation (supply voltage, temperature, electromagnetic environment, aging, technological dispersion).

The algorithms that we have implemented not only output the statistical parameters (α, μ, σ^2) of a OJMD but also curves providing information about the quality of the measures. Denote by $T_i(t)$ (resp. $T_{ech}(t)$) the period of O_i (resp. O_{ech}) at time t. Denote by (b_j^i) the output bit sequence of OJMD_i where the bit b_j^i is sampled at time t_j. One can compute a good approximation of $T_{ech}(t_{j_0})/T_i(t_{j_0}) \mod 1$ using Fact 2 by computing $\frac{1}{2}\mathbb{P}_{S_{j_0}}\{b_i \neq b_{i+1}\}$ with $S_{j_0} = \{b_{j_0}^i, \ldots, b_{j_0+L}^i\}$ for L big enough so that one can compute the empirical probability $\mathbb{P}_{S_{j_0}}\{b_i \neq b_{i+1}\}$ is small enough so that period of time between $b_{j_0}^i$ and $b_{j_0+L}^i$ is short.

5.2 Transient phenomena

In Figure 7, we see the evolution of $T_{ech}(t)/T_0(t) \mod 1$ when the acquisition of data begins just when O_{ech} and O_0 are enabled. We note that $T_{ech}(t)/T_0(t) \mod 1$ gradually slides down to a stable value. Our interpretation of the transient phenomenon is that when enabled the ring produce heat which increase the temperature locally on the silicon and thus modify the signal propagation time. This means that it is important to let some time passes between when we enable the OJMD and the acquisition time to let it stabilize. If one does not take this precaution, the variation of $T_{ech}(t)/T_0(t) \mod 1$ can distort the measuring result and increase the estimation of the physical noise.

5.3 Experiments and results

The method of continued fraction of Section 4.2 is used in order to compute a value of N well suited so that $S_{N,M}$ gives a good approximation of the probability density function $D(M)$ for each acquisition file (Figure 8 for the OJMD_0). We compute $V(S_{N,M})$ which

index	nb_inv	d_ro_min ns	d_ro_max ns	f_max mhz	f_min mhz
ech	15	4,993	14,677	100,14	34,07
0	16	5,377	15,821	92,99	31,6
1	17	5,54	16,344	90,25	30,59
2	18	5,944	17,508	84,12	28,56
3	19	6,336	18,683	78,91	26,76
4	20	6,74	19,847	74,18	25,19
5	21	7,211	21,159	69,34	23,63
6	22	7,591	22,275	65,87	22,45
7	23	7,779	22,958	64,28	21,78
8	24	8,183	24,122	61,1	20,73
9	25	8,346	24,645	59,91	20,29
10	26	8,749	25,808	57,15	19,37
11	27	9,057	26,389	55,21	18,95
12	28	9,541	27,889	52,41	17,93
13	29	9,915	28,981	50,43	17,25
14	30	10,349	30,288	48,31	16,51
15	31	10,632	31,067	47,03	16,09
16	16	5,472	15,913	91,37	31,42
17	17	5,65	16,411	88,5	30,47
18	18	6,083	17,717	82,2	28,22
19	19	6,292	18,322	79,47	27,29
20	20	6,725	19,628	74,35	25,47
21	21	7,1	20,721	70,42	24,13
22	22	7,413	21,672	67,45	23,07
23	23	7,726	22,743	64,72	21,98
24	24	8,129	23,904	61,51	20,92
25	25	8,341	24,471	59,94	20,43
26	26	8,733	25,621	57,25	19,52
27	27	9,227	26,901	54,19	18,59
28	28	9,489	27,921	52,69	17,91
29	29	9,946	29,212	50,27	17,12
30	30	10,349	30,373	48,31	16,46
31	31	10,539	30,997	47,44	16,13

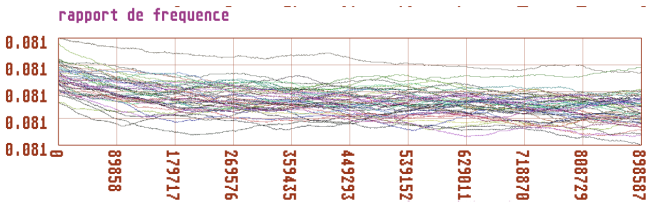

Figure 7: *Evolution of* $T_{ech}(t)/T_0(t) \mod 1$ *at startup*

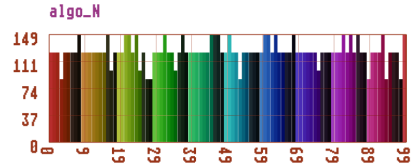

Figure 8: $S_{N,M}$ *for* $OJMD_0$.

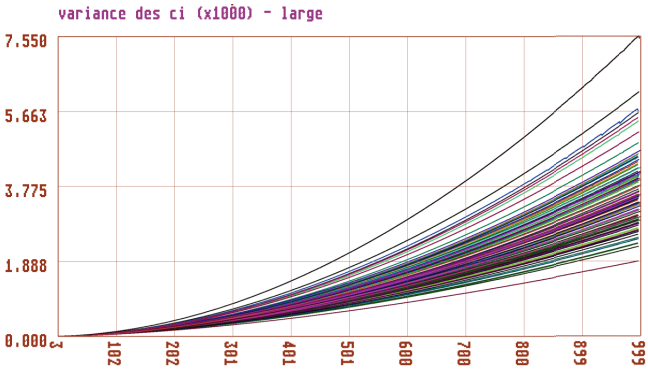

Figure 9: $V(S_{N,M})$ *as a function of* M *for a large span of* M.

are normalized by $2.N$ so that they are in the interval $[0, 0.5]$. The Figure 9 shows $V(S_{N,M})$ as a function of M for a large span of M (from 0 to 1000). We note the quadratic shape of the curve due to flicker noises when M is big. This quadratic behavior is not visible on the simulated ring with only thermal noise.

In order to compute the slope of thermal component of the curve $V(S_{N,M})$, we use a span of accumulation time M where the effect of flicker noise seems to be negligible: with the condition $M \gg N$, we can chose M in the interval $[500, 700]$ with our settings. The Figure 10 represents the values of $V(S_{N,M})$ for M in the span $[500, 700]$. Only the squared points correspond to computed values and the lines between these points are interpolations. We remark that in this span the law of $V(S_{N,M})$ is nearly affine. The Figure 11 gives the centered distribution of the c_i for $M = 700$. The general aspect of the law is Gaussian as expected and we not see any hint of a folding that we experienced in the simulations (see Section 3). We have furtherly checked this assumption by computing successive moments of the law and comparing them with that of a Gaussian law. The Figure 12 provides with the histogram of the values of σ^2 that is the slope of the affine law $V(S_{N,M})$ computed by linear interpolation. We note that the log normal distribution is Gaussian probably because of the flicker noise. We this dataset we obtain a most values of Q are around $2.16\,10^{-6}$.

The Figure 13 gives the histogram of the values of Q the quality factor of a simulated OJMD. With this data set, we obtain a value of Q of $2.36\,10^{-6}$ for a Q injected in simulation of $2.39\,10^{-6}$.

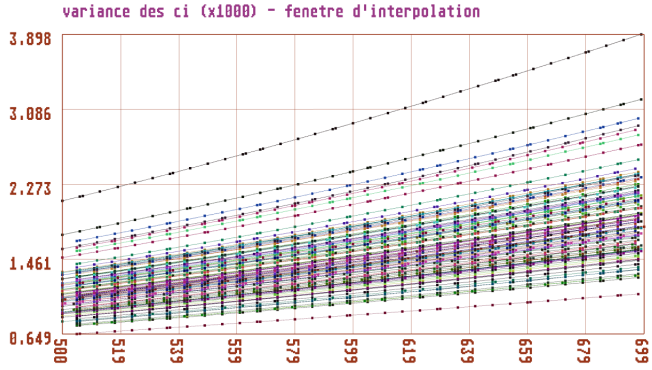

Figure 10: $V(S_{N,M})$ *as a function of* M *for* $M \in [500, 700]$.

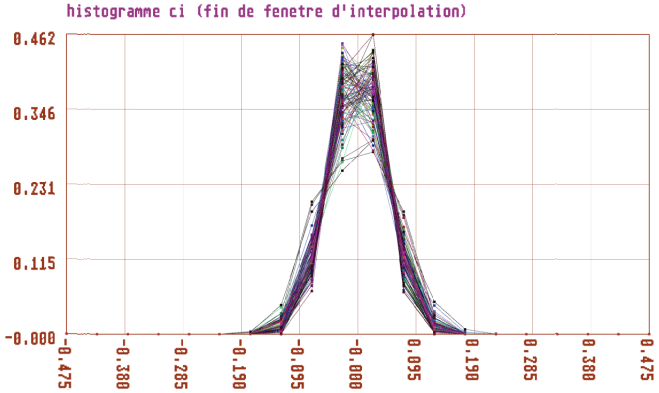

Figure 11: *Distribution of* c_i *for* $M = 700$.

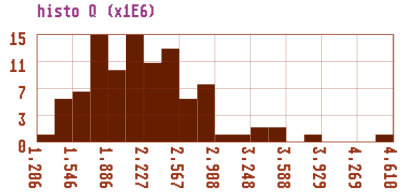

Figure 12: *Histrogram of values of* Q *for simulated dataset.*

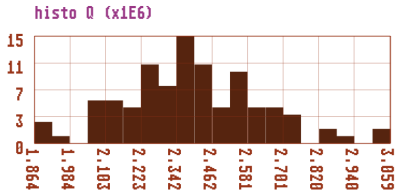

Figure 13: *Histrogram of values of Q.*

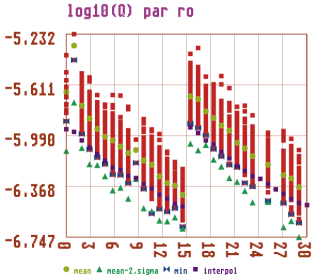

Figure 14: *Values of Q for different OJMD.*

In Figure 14, we see the values of Q for different OJMD. We find as expected that the value of Q is increasing with the frequency of the sampled oscillator.

The Figure 15 gives the value of T_i and T_{ech} for each OJMD. Some OJMD could not be characterized because the continued fraction method could not allow us to obtain a valid value of N.

6 Conclusion

In this paper, we have shown that the jitter measurement method of [] works at the condition that one choose correctly the two parameters M and N upon which it depends. Using simulations, we have highlighting the fact that bad choices of N and M produces

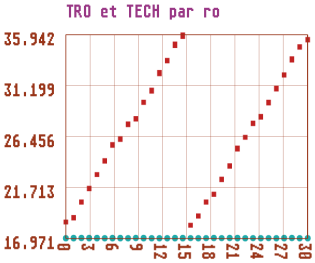

Figure 15: *Values of T_i and T_{ech} for each OJMD.*

outliers which affect the measure of the σ^2 parameter of the phase jitter. We have presented method that allows to choose correctly M and N which where not explained in [2] and tested the resulting algorithms with simulations.

References

[1] W. Schindler and W. Killmann. Evaluation Criteria for True (Physical) Random Number Generators Used in Cryptographic Applications. In B. S. Kaliski, C. Koç, and C. Paar, editors, *Cryptographic Hardware and Embedded Systems – CHES 2002*, volume 2523 of *LNCS*, pages 431–449. Springer, 2003.

[2] V. Fischer and D. Lubicz. Embedded Evaluation of Randomness in Oscillator Based Elementary TRNG. In L. Batina and M. Robshaw, editors, *Cryptographic Hardware and Embedded Systems – CHES 2014*, volume 8731 of *LNCS*, pages 527–543. Springer, 2014.

[3] M. Baudet, D. Lubicz, J. Micolod, and A. Tassiaux. On the Security of Oscillator-Based Random Number Generators. *Journal of Cryptology*, 24:398–425, 2011.

[4] V Fischer and D. Lubicz. Entropy computation for oscillator based trng.

[5] D. Lubicz and N. Bochard. Towards an Oscillator Based TRNG with a Certified Entropy Rate. *IEEE transaction on computers*, 63:1–10, 2014.

[6] A. Ya. Khinchin. *Continued fractions*. The University of Chicago Press, Chicago, Ill.-London, 1964.