Max Bias Analysis:

A New Approach on Computing the Entropy of Free Ring-Oscillator

Nicolas David and Eric Garrido

Thales, Gennevilliers, France, nicolas-i.david@thalesgroup.com
eric.garrido@thalesgroup.com

Abstract. This work introduce a new approach called *Max bias analysis* for the entropy computation of structures of Free Ring Oscillator-based Physical Random Number Generator. It employs the stochastic model based on the well-established Wiener process, specifically adapted to only capture thermal noise contributions while accounting for potential non-zero bias in the duty cycle. Our analysis is versatile, applicable to combinations of multiple sampled Ring Oscillator (RO) filtering by any function. The entropy computation takes as inputs the parameters of the thermal stochastic model and delivers directly a proven bound for both Shannon entropy and min-entropy to fulfill AIS31 and NIST SP 800-90 B. As an example, we apply the new methodology on an enhanced structure of TRNG combining several free-running Ring Oscillators filtered by a vectorial function built from a linear error correcting code that optimizes the functional performance in terms of [entropy rate/silicium area used] and that maintains the mathematical proof of the entropy lower bound as simple as possible.

Keywords: Ring Oscillator · MO-TRNG · Thermal Noise · Entropy · Bias · Hidden Wiener Process · Vectorial Conditioning

1 Introduction

True Random Number Generators (TRNGs) are indispensable for ensuring the unpredictability of cryptographic secrets such as encryption keys, nonces, and random masks in countermeasures against side-channel attacks. TRNGs leverage physical sources of randomness to ensure unpredictability at their core. However, the quality and reliability of TRNGs depend critically on the entropy of the underlying physical noise sources.

Ring Oscillators (ROs) are one of the most studied and widely used physical sources of randomness in TRNGs ([BLMT11], [KG04], [Saa21],[VD10], [HTBF14]), their randomness is derived from noise sources such as *thermal noise* and *flicker noise*. If recent works [BCF+24, PV24, Sko24] have shown that the impact of flicker noise could be used to enhance the entropy bounds, its early autocorrelation leads to complicated modeling. On the opposite, thermal noise, originating from the random motion of charge carriers, is an uncorrelated source of randomness and can be modeled with a Wiener process as shown in [BLMT11]. By analogy to the hidden Markov process, in which the observable random variable $\{Y_n\}_{n\in\mathbb{N}}$ can be seen as a noisy or a partial observation of a Markov chain $\{X_n\}_{n\in\mathbb{N}}$, the output sequence reveals only partial information about the underlying phase. This leads to the concept of a *Hidden Wiener Process*.

Entropy of MO-TRNG. Along with their modeling, the authors of [BLMT11] provided formulas for probability and entropy only for the special case of an unbiased duty cycle (inherent parameter of a ring oscillator). Building upon this foundational work, subsequent studies [LF24, Saa21] have developed methodologies to compute bounds on entropy of the

single or the Multi Ring Oscillator True Random Number Generator (MO-TRNG). Both of the developed approaches are numerical and can be interpreted as a Hidden Markov process approximation of the Wiener process. This technique exhibits the following limitations:

- **Theoretical Confidence.** The Wiener to Markov approximation is a plug-in technique, thus derived bounds that are ad-hoc in nature, lacking a formal proven lower bound on entropy.

- **Computation Cost.** The entropy estimation requires numerical computations since there is no close form for the entropy of a hidden Markov process. When analyzing structure with multiple ring oscillators, this high computational demands results in a high complexity in numerical applications.

- **Design Efficiency.** Previous works have focused their analysis on the natural filtering function: the boolean XOR. However, restricting the output of the MO-TRNG structure to a single bit can result in a loss of entropy. Indeed, when sampling L independent ROs, the entropy rate prior to the filtering function might be high, in particular greater than 1. However, the rate of entropy of the output stays smaller than 1 for all the selected sampling periods, when employing boolean filtering. By allowing the TRNG to generate multiple random bits per sample, we can explore more complex filtering functions that can enhance the bitrate optimization.

Our proposed methodology, *Max Bias Analysis*, addresses all these deficiencies effectively by providing a simpler approach to compute entropy bounds, hence allowing the derivation of formulas for biased duty cycles and proven lower bounds for MO-TRNG structures with non-trivial vectorial post-processing.

Contributions. This research contributes a novel methodology building upon prior work by [BLMT11]. The key contributions are as follows:

- We introduce mathematical formulas to determine the probability of a sampled sequence as well as entropy bounds that are consistent with existing methodologies for the case of an unbiased duty cycle and remain valid in scenarios with low jitter levels (high sampling frequency) and biased duty cycles.

- We provide a proven lower bound for the rate of entropy that can be directly computed, even in the context of complex multi-ring oscillator configurations encompassing multiple ring oscillators and vectorial combiners.

- We conduct a comprehensive analysis of the filtering function, aimed at optimizing the performance of the MO-True Random Number Generator (TRNG) for both Shannon and min-entropy purposes. The best results are obtained while considering a linear vectorial conditioner; indeed, we obtain an increase of the output bitrate with respect to the XOR by a factor of 4 for the case of 32 ring oscillators and almost 10 in the case of 128 ring oscillators.

Organization. This paper organizes as follows: in Section 2, we introduce some notations and mathematical tools; Section 3 presents the Wiener stochastic modelization of [BLMT11]; then in Section 4, we introduce the max bias analysis in the case of a single ring oscillator, deriving a formula for the probability of a sampled sequence; the associated entropy is studied in Section 5; finally in Section 6, we study the case MO-TRNG, in particular we present vectorial conditioners that significantly enhance the output bitrate.

Table 1: Summary of Notations

X	Random variable over $\{0,1\}^r$	
p	Probability density function of a random variable X	$p(x) = \mathbb{P}(X = x)$
e	Relative bias of a random variable X	$e(x) = 2^r\mathbb{P}(X = x) - 1$
ϵ	Linear Bias	$\epsilon(u) = \mathbb{E}\left((-1)^{u\cdot X}\right)$
$H(X)$	Shannon entropy of X	
$H_\infty(X)$	Min-entropy of X	
C	Moment of order 2	Theorem 1
Δ	Residue function	Theorem 1
$h(B)$	Entropy of a Bernoulli of relative Bias B	Proposition 4
RO	Ring Oscillator	
ΔT	Sampling period	
ϕ	Phase of a Ring Oscillator	
ω, σ	Mean and standard deviation of the Wiener Process	
Q	Quality factor	$Q = \sigma^2\Delta T$
$\mathcal{S}$	Signal function	
α, a	Duty cycle and relative duty cycle of a RO	$a = 2\alpha - 1$
$(s(n))_{n\in\mathbb{N}}$	Sampled sequence	
$\ggg_r$	Translation of r bits to the right	
F	Filtering function of the MO-TRNG	
$(f(n))_{n\in\mathbb{N}}$	Output sequence of the MO-TRNG	
$B(\alpha, Q)$	Max Bias of a RO	
τ_S^C	Conditional Shannon entropy rate	
τ_∞^C	Conditional min-entropy rate	
$\hat{F}$	Walsh transform of F	

2 Mathematical tools and Notations

In this section, we will introduce some notations and mathematical tools useful to model ring oscillators and compute entropy bounds of their sampled sequence.

2.1 Walsh Transform

First, we define the Walsh transform, a well-known tool that is a specific case of the discrete Fourier transform.

Definition 1. Consider a function $F : \{0,1\}^L \to \mathbb{R}$, we call Walsh transform of F the function:

$$\begin{array}{rccc} \hat{F}: & \{0,1\}^L & \to \mathbb{R} \\ & w & \mapsto \sum_{x\in\{0,1\}^L}(-1)^{w\cdot x}F(x) \end{array},$$

that is reverse by the following operation:

$$F(x) = \frac{1}{2^L}\sum_{w\in\{0,1\}^L}(-1)^{x\cdot w}\hat{F}(w)$$

For Boolean function $F : \{0,1\}^L \to \{0,1\}$, it is often preferred to study the function $(-1)^F$ rather than F. To ease notations, we will denote by $\hat{F}$ the Walsh transform of $(-1)^F$:

$$\begin{array}{rccc} \hat{F}: & \{0,1\}^L & \to \mathbb{R} \\ & w & \mapsto \sum_{x\in\{0,1\}^L}(-1)^{w\cdot x+F(x)} \end{array}$$

Following this approach, we will define the Walsh transform of vectorial function $F : \{0,1\}^L \to \{0,1\}^r$ by the Walsh transforms of $(-1)^{u\cdot F}$ for each nonzero u. Hence,

$$\begin{array}{rccl} \hat{F}: & \{0,1\}^L \times \{0,1\}^r \setminus \{0\} & \to & \mathbb{R} \\ & (u,w) & \mapsto & \sum_{x\in\{0,1\}^L} (-1)^{x\cdot x + u\cdot F(x)}. \end{array}$$

2.2 Discrete Random Variable

The primary focus of our study is a discrete random variable X, which takes values in $\{0,1\}^r$ for some $r \geq 1$. There are two equivalent approaches to defining its probability:

- **Density function.** This is a straightforward approach where we consider the function: $p : x \mapsto \mathbb{P}(X = x)$. Alternatively, it is possible to consider the point-wise distance to the uniform distribution. Hence, defining the relative bias:

  $$e(x) = 2^r \mathbb{P}(X = x) - 1,$$

  equivalently, one could write:

  $$p(x) = \frac{1}{2^r}(1 + e(x)).$$

- **Linear Bias.** Alternatively, a random variable can be defined by its linear bias. More precisely, it consists of the following function, defined for all $u \in \{0,1\}^r \setminus \{0\}$:

  $$\epsilon(u) = \mathbb{E}\left((-1)^{u\cdot X}\right).$$

Now, we will share some properties that connects relative bias, linear bias and probability function.

Proposition 1 (link between relative and linear bias)**.** *Let p and e be the function defined as before. The following equations are verified:*

$$e(x) = \sum_{u\neq 0} (-1)^{u\cdot x} \epsilon(u), \qquad \begin{aligned} \epsilon(u) &= \sum_x p(x)(-1)^{u\cdot x} \\ &= \frac{1}{2^r} \sum_x e(x)(-1)^{u\cdot x}. \end{aligned}$$

The proof of these equations directly follows the definitions of the Walsh and the reverse Walsh transforms. These equations confirm the equivalence of both approaches to defining a random variable, and they enable us to establish certain bounds, as outlined in Corollary 1.

Corollary 1. *Let p and e be the function defined as before. If there exists a bound B such that for all $u \in \{0,1\}^r \setminus 0$, $\mid \epsilon(u) \mid \leq B$ then*

$$\forall x \in \{0,1\}^r, \mid e(x) \mid \leq (2^r - 1)B.$$

Similarly, if there exists a bound B such that for all $x \in \{0,1\}^r$, $\mid e(x) \mid \leq B$ then

$$\forall u \in \{0,1\}^r \setminus 0, \mid \epsilon(u) \mid \leq B.$$

Remark 1. In the case where $r = 1$, Corollary 1 becomes degenerate since e and ϵ coincide. More broadly, the scenario $r = 1$ is unique and will be analyzed separately throughout this work.

2.3 Some Properties

In this section, we will share some well-known properties based on the Walsh transform: the Parseval Identity and the Pilling up Lemma. The first one connects moments of order 2 of the Walsh transform and relative bias; the latter describes the behavior of the Walsh transform of a tensor product.

Proposition 2 (Parseval identity [Ste04])**.** *Let p and e be the function defined as before. The following equation is verified:*

$$\sum_{x\neq 0} \epsilon(x)^2 = \frac{1}{2^r}\sum_x e(x)^2$$

This property follows the orthogonality of the Walsh Basis.

Proposition 3 (Pilling up Lemma [Mat93])**.** *Consider $(X_i)_{1\leq i\leq n}$ as n independent discrete random variables such that X_i takes its values in $\{0,1\}^{r_i}$. For each index i, we denote by p_i the probability function of X_i. If we consider the random variable $X = (X_i)_{1\leq i\leq n}$ as a discrete random variable on $\{0,1\}^{\sum_i r_i}$ and denote by p the probability function, the following equations are verified:*

$$\epsilon(u) = \mathbb{E}((-1)^{uX}) = \mathbb{E}((-1)^{\sum_i u_i X_i}) = \prod_i \mathbb{E}((-1)^{u_i X_i}) = \prod_i \epsilon_i(u_i).$$

2.4 Entropy of random variable on $\{0,1\}^r$

In this section, we will study another probabilistic object called entropy.

Definition 2. Let X be a random variable on $\{0,1\}^r$, we call entropy of X the following quantity:

$$H(X) = \sum_{x\in\{0,1\}^r} -\log_2(p(x))\cdot p(x)$$

Let's note that it is possible to rewrite the definition of entropy in function of e by replacing p, hence we obtain :

$$H(X) = \frac{1}{2^r}\sum_{x\in\{0,1\}^r} -\log_2(\frac{1+e(x)}{2^r})\cdot(1+e(x))$$

To the best of our knowledge, this result presented in the next theorem is new and provides a bond of the entropy from a bond on the relative bias.

Theorem 1. *In this theorem, we consider X to be a random variable on $\{0,1\}^r$ and present bounds on its entropy based on analyzing the relative or linear bias using the following functions:*

$$C(X) = \frac{1}{2^r}\sum_x e(x)^2 = \sum_{x\neq 0}\epsilon(x)^2,$$
$$\Delta(B) = \frac{1}{\ln(2)}\left[(1-B)\ln(1-B) + B - B^2/2\right].$$

- ***Relative Bias*** *Let $B_e \in \mathbb{R}_+$ be a bound on the relative bias, i.e., $\forall x \in \{0,1\}^r$, $\mid e(x) \mid \leq B_e$. If $B_e \leq 1$, then the following bound on the entropy is verified:*

$$H(X) \geq r - \frac{C(X)}{2\ln 2} - \Delta(B_e)$$

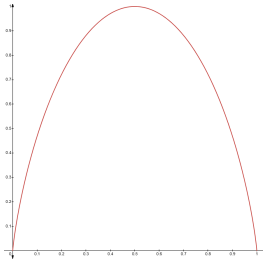

Figure 1: Plot of $p \mapsto -p \log p - (1-p)\log(1-p)$ representing the Bernoulli of probability p. It is increasing for $p \in [0, 1/2]$ and decreasing for $p \in [1/2, 1]$.

- ***Linear Bias.*** *Let $B_\epsilon \in \mathbb{R}_+$ be a bound on the linear bias, i.e $\forall x \in \{0,1\}^r$, $\mid \epsilon(x) \mid \leq B_\epsilon$. If $B_\epsilon \leq \frac{1}{2^r - 1}$, then the following bound on the entropy is verified:*

$$H(X) \geq r - \frac{C(X)}{2 \ln 2} - \Delta((2^r - 1)B_\epsilon)$$

The proof of this theorem is presented in Appendix A, it consists of developing the log function in a power series and studying the different moments of X.

As advertised earlier, the specific case $r = 1$ allows the derivation of a tighter lower bound for the entropy of $H(X)$.

Proposition 4. *Let X be a random variable on $\{0,1\}$, and $B \in [0,1]$ be a bound on the relative or the linear bias:*

$$\forall x \in \{0,1\}^r, \mid \epsilon(x) \mid \leq B.$$

The following bound on the entropy is verified:

$$H(X) \geq h(B) = 1 - \frac{(1+B)\log_2(1+B)}{2} - \frac{(1-B)\log_2(1-B)}{2}.$$

The proof of this proposition consists of studying Figure 1. We remark that when the relative bias increases, the entropy decreases. Therefore, the entropy of X is bounded by the entropy of the Bernoulli of relative bias B.

3 Stochastic Modelization with Hidden Wiener Process

Research on modern physical RNG entropy estimation was first conducted by Killman and Schindler [KS08], whose stochastic model employs independent and identically distributed transitions of time (half-periods) to represent jitter. Later [BLMT11] provided a stochastic modelization of the free MO-TRNG using Wiener process. In this section, we recall this modelization following the depiction done in Figure 2. This process involves different steps: *the phase, the signal, and the sampled sequence.*

Phase. The phase ϕ of a ring oscillator RO can be effectively modeled using a Wiener process, which accounts for the jitter introduced by thermal noise. Note that the thermal noise is not the only contributor to the jitter. Indeed, recent studies have been working on the impact of the flicker noise [BCF+24]. We will nonetheless focus on thermal noise since there is a lot of confidence in its modelization, and any additional noise should only increase the entropy. For the thermal modelization, there exist parameters such as a drift ω and a deviation σ such that for any time delay ΔT, the following relationship holds:

$$\phi(t + \Delta T) - \phi(t) \sim \mathcal{N}(\omega \Delta T, \sigma^2 \Delta T),$$

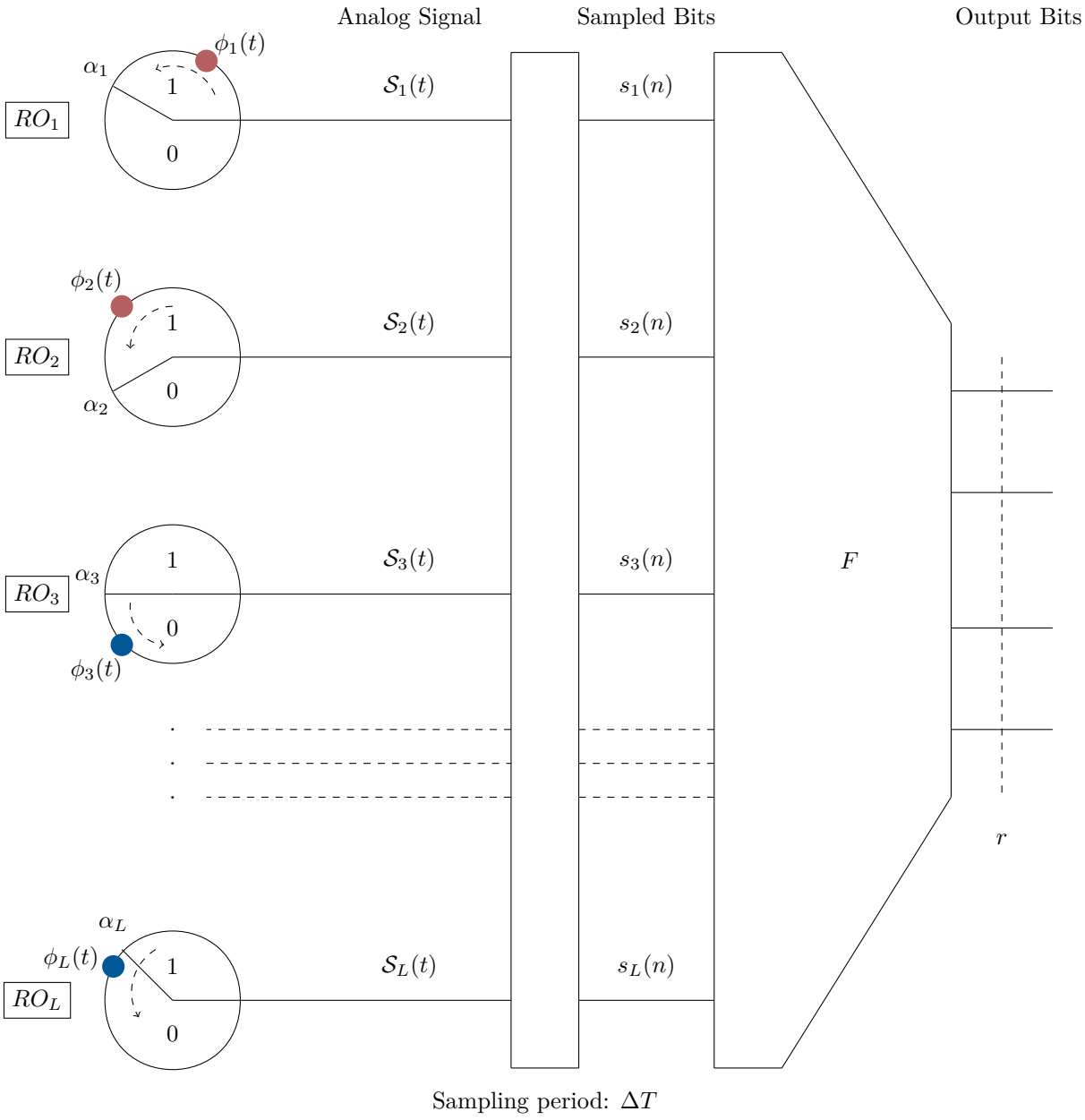

Figure 2: Multi Free Ring Oscillator Physical Random Number Generator modelization.

where $\mathcal{N}(\omega\Delta T, \sigma^2\Delta T)$ stands for a Gaussian distribution of mean $\omega\Delta T$ (deterministic component determined by the mean period of the RO) and variance $\sigma^2\Delta T$ (thermal jitter component). Regarding the initial phase at time t_0 denoted by ϕ_0, this work focuses on studying the process that begins with a uniform initial distribution, formally expressed as $\phi_0 \mod 1 \sim \mathbb{U}$. This choice is justified by both practical and theoretical reasons. Indeed, the entropy rate defined as a limit hereafter is independent from the initial distribution of the phase and corresponds to the one obtained directly starting with the stationary distribution. Furthermore, allowing the phase to drift over an extended period leads to sufficient accumulated jitter, which can support the assertion that $\phi_0 \mod 1 \sim \mathbb{U}$. Some specific consequences of this choice for the initial phase will be later discussed in Section 4.

Analog Signal. The signal generated by an oscillator is simplified as a clock signal that alternates between two values. A parameter α (resp. $a = 2\alpha - 1$) called *duty cycle* (resp. *relative duty cycle*) models the difference in time spent by the signal between the two values. Hence, a visual representation of this signal can be depicted by observing the evolution of $\phi(t)$ reduced modulo 1, effectively mapping it onto a circle of length 1, divided into two distinct regions: *area 1* and *area 0*. More precisely, the signal function $\mathcal{S}(t)$ can be defined as follows:

$$\begin{array}{rccl} \mathcal{S}: & \mathbb{R} \to & \{0,1\} & \\ & t \mapsto & \begin{cases} 1 & \text{if } \phi(t) \mod 1 \in [0,\alpha[\\ 0 & \text{else} \end{cases} & \end{array}$$

(Discrete) Sampled sequence. From an analogical signal $\mathcal{S}$, its measurement at regular intervals yields a binary sampled sequence. Hence, for a fixed delay ΔT, we define the sequence $(s(n))_{n\in\mathbb{N}}$ by

$$s(n) = \mathcal{S}(t_0 + n\Delta T) \in \{0,1\}.$$

From an observer's point of view, the sampled sequence is the only element that can be witnessed. The associated phase stays hidden, thus following what we call a *hidden Wiener process*. We denote by $(\alpha, \omega, \sigma, \Delta T) - RO$ this whole process that results in the production of the sequence $s(n)$.

MO-TRNG. As illustrated in Figure 2, a MO-TRNG is build from a sequence of L ring oscillators $((\alpha_i, \omega_i, \sigma_i, \Delta T) - RO_i)_{1\leq i\leq L}$. For each ring oscillator, we build sampled sequences using a uniform time interval ΔT, resulting in a collection of sampled sequences: $(s_i(n))_{1\leq i\leq L, n\in\mathbb{N}}$. This collection is subsequently processed through a filtering function $F : \{0,1\}^L \to \{0,1\}^r$producing the output, which is the TRNG sequence:

$$f(n) = F(s_1(n), \cdots, s_L(n)) \in \{0,1\}^r.$$

Remark 2. Besides the sampling time ΔT, all the parameters (duty cycle, standard deviation, drift) depend on the ring oscillator. Hence, while studying RO_i, we will denote α_i as its duty cycle, ω_i as the drift, and σ_i as the standard deviation of ϕ_i.

4 Study of the Sampled sequence Starting with Uniform Distribution

In this section, we will study the case of a single ring oscillator and leverage the hypothesis regarding the initial distribution to derive results on the probability of sampled sequences. First, we will demonstrate results that can be interpreted as temporal symmetry of the sampled sequences, then we will define a tool useful to provide bounds throughout this work: *max bias*, and finally we will put together all these ideas to derive a formula for the density probability function of a sampled sequence.

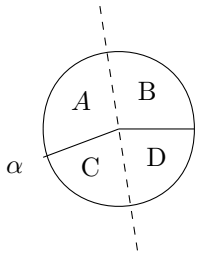

Figure 3: Symmetry of a RO of parameter α required for ϕ_0.

4.1 Temporal Symmetries

In the following, we will present some results on the symmetry of the sampled sequence in the case of a uniform initial phase.

Reversing Time. The main idea to highlight is that time can be considered in both directions. In that regard, knowing the phase at time t: $\phi(t) = \phi_t$, we will define the two Wiener processes W^+ and W^- that define time jumps forwards and backwards, respectively:

$$\underset{t \to t+\Delta T}{W^+} : \phi(t + \Delta T) = \phi_t + \mathcal{N}(\omega \Delta T, \sigma^2 \Delta T)$$
$$\underset{t \to t-\Delta T}{W^-} : \phi(t - \Delta T) = \phi_t + \mathcal{N}(-\omega \Delta T, \sigma^2 \Delta T)$$

Following these Wiener processes, we denote by $\mathcal{S}^+$ and s^+ (resp. $\mathcal{S}^-$ and s^-) the signal function and the sampled sequence obtained with W^+ (resp. W^-) starting with ϕ_0 at time t_0. The following lemmas will present some link between these objects.

Lemma 1. *If the initial distribution ϕ_0 is symmetric with respect to the angle bisector of α (as depicted in Figure 3), then the signal functions $\mathcal{S}^+$ and $\mathcal{S}^-$ share the same probability distribution. More precisely, if we denote by $\mathcal{R}_\alpha$ the reflection along the angle bisector:*

$$[\phi_0 = R_\alpha(\phi_0)] \Rightarrow [\mathcal{S}(t_0 + t) \sim \mathcal{S}(t_0 - t)]$$

This lemma consists of studying the action of R_α on the different parts that compose the signal function.

- The initial distribution ϕ_0 is unchanged by hypothesis.
- The time t is reversed, hence turns a W^+ execution into a W^- execution.
- The measurement areas (corresponding to 0 and 1) are unchanged. Indeed, as depicted in Figure 3, areas A and B (resp. C and D) are swapped.

If the symmetrical aspect of $\mathbb{U}$ allows for the derivation of Lemma 1, its stationary aspect with respect to the Wiener results in Lemma 2.

Lemma 2. *If the initial phase ϕ_0 follows a uniform distribution, then for any time t, $\phi(t)$ follows a uniform distribution. As a consequence, the signal function is unchanged by any temporal shift:*

$$[\phi_0 \sim \mathbb{U}] \Rightarrow [\mathcal{S}(t_0 + t) \sim \mathcal{S}(t'_0 + t)]$$

This is a consequence of the stationary aspect of the uniform distribution through Wiener evolution. Hence, for any time t, $\phi(t) \sim \mathbb{U}$. Therefore, an execution defined by W^+ starting at time t_0 will be indistinguishable from an execution starting at time t.

Theorem 2 (The probability function self-reverses.)**.** *Consider an* $(\alpha, \omega, \sigma, \Delta T) - RO$ *and a sequence of* n *boolean* $(b_i)_{1 \leq i \leq n} \in \{0,1\}^n$*, then the following equation holds:*

$$\mathbb{P}\left[\forall i,\ s(i) = b_i\right] = \mathbb{P}\left[\forall i,\ s(n-i) = b_i\right]$$

Proof. This proof consists of gathering the ideas presented both in Lemma 1 and Lemma 2 together. Hence,

$$\begin{aligned}
\mathbb{P}\left[\forall i,\ s(i) = b_i\right] &= \mathbb{P}\left[\forall i,\ \mathcal{S}(t_0 + i\Delta T) = b_i\right] \\
&\underset{Lemma\ 1}{=} \mathbb{P}\left[\forall i,\ \mathcal{S}(t_0 - i\Delta T) = b_i\right] \\
&\underset{Lemma\ 2}{=} \mathbb{P}\left[\forall i,\ \mathcal{S}((t_0 + n\Delta T) - i\Delta T) = b_i\right] \\
&= \mathbb{P}\left[\forall i,\ \mathcal{S}(t_0 + (n-i)\Delta T) = b_i\right] \\
&= \mathbb{P}\left[\forall i,\ s(n-i) = b_i\right]
\end{aligned}$$

□

This theorem allows identifying sequences that share the same probability of being sampled. Based on this fact, we are able to derive an exact formula for the probability of a sequence (Subsection 4.3). However, the formula obtained is not practical; hence, we introduce *max bias* to derive a practical approximation of the probability density function.

4.2 Max bias Analysis

While previous research [BLMT11] has yielded theoretical results concerning the probability in the particular case where $\alpha = 1/2$, this study aims to extend these findings to encompass all values of α. However, up to a symmetric transformation that interchanges areas 0 and 1, we can reasonably assume that $\alpha \geq 1/2$. Therefore, for the remainder of this study, we will adopt this assumption. This section aims at introducing a tool for providing efficient bounds: *the max bias function.*

Definition 3 (Max bias)**.** For a given $(\alpha, \omega, \sigma, \Delta T) - RO$, we define the max bias as the following quantity:

$$B_{\omega,\sigma}(\alpha, \Delta T) = \max_{x \in [0,1]} \left| \mathbb{E}[(-1)^{\mathcal{S}(t_0 + \Delta T)} | \phi(t_0) = x] \right|$$

Remark 3. The max bias asymptotical behavior is as follows: $\lim_{\Delta T \to 0} B_{\omega,\sigma}(\alpha, \Delta T) = 1$ and $\lim_{\Delta T \to \infty} B_{\omega,\sigma}(\alpha, \Delta T) = a$.

Computing Max Bias. We will now outline a method for computing the maximum bias. First, Lemma 3 identifies the point x that maximizes $\left|\mathbb{E}[(-1)^{\mathcal{S}(t_0+\Delta T)} | \phi(t_0) = x]\right|$. Building on this result, Theorem 3 will present an explicit formula for the maximum bias. Finally, we will check the consistency of our result by confronting it with formulas presented in [BLMT11] in the specific case where $\alpha = 1/2$.

Lemma 3 (Middle point trick)**.** *For a given sampling time* ΔT*, the bias* $\left|\mathbb{E}[(-1)^{\mathcal{S}(t_0+\Delta T)} | \phi(t_0) = x]\right|$ *is maximized for:*

$$x^* = \alpha/2 - \omega\Delta T$$

As a consequence, $B(\Delta T) = \left|\mathbb{E}[(-1)^{\mathcal{S}(t_0+\Delta T)} | \phi(t_0) = x^*]\right|$

As depicted in Figure 4, the bias is maximized when the jitter Gaussian is centered in the middle of the largest area. Since we suppose $\alpha \geq 1/2$, then the 0 area is larger. By taking into account the mean of the Gaussian from the Wiener process, the max bias is obtained when $x^* = \alpha/2 - \omega\Delta T$.

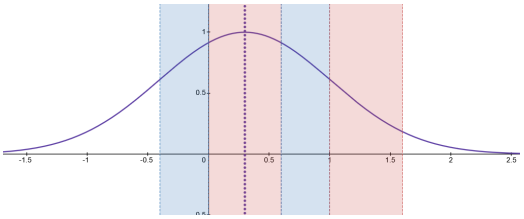

Figure 4: Maximizing the bias when $\phi_0 = \alpha/2 - \omega\Delta T$. The mean of the Gaussian is aligned with $\alpha/2$, the 0 area is depicted in red, and the 1 area is depicted in blue.

Table 2: $B(\alpha, Q)$ in function of Q and α.

$a = 2\alpha - 1$	0	0.001	0.01	0.05	0.1	0.2
$Q = \sigma^2\Delta T$						
0.01	0.9751613	0.9753361	0.977	0.983	0.9881	0.9946
0.02	0.8458005	0.8463909	0.852	0.873	0.8963	0.9322
0.03	0.7022005	0.7030126	0.71	0.741	0.7754	0.8336
0.04	0.5777546	0.5786689	0.587	0.622	0.6623	0.7337
0.05	0.4744875	0.4754483	0.484	0.521	0.5649	0.6441
0.06	0.3895303	0.3905123	0.399	0.437	0.483	0.5672
0.07	0.3197597	0.3207514	0.33	0.368	0.415	0.5026
0.08	0.2624819	0.2634779	0.272	0.311	0.3589	0.449
0.09	0.2154633	0.2164614	0.225	0.265	0.3126	0.4046
0.1	0.1768671	0.1778662	0.187	0.226	0.2746	0.3681
0.2	0.0245688	0.0255688	0.035	0.074	0.1243	0.2234
0.3	0.0034129	0.0044129	0.013	0.053	0.1034	0.2032
0.4	0.0004741	0.0014741	0.01	0.05	0.1005	0.2005
0.5	$6.59 \cdot 10^{-5}$	0.0010659	0.01	0.05	0.1001	0.2001
0.6	$9.15 \cdot 10^{-6}$	0.0010091	0.01	0.05	0.1	0.2
0.7	$1.27 \cdot 10^{-6}$	0.0010013	0.01	0.05	0.1	0.2
0.8	$1.77 \cdot 10^{-7}$	0.0010002	0.01	0.05	0.1	0.2
0.9	$2.45 \cdot 10^{-8}$	0.001	0.01	0.05	0.1	0.2
1	$3.41 \cdot 10^{-9}$	0.001	0.01	0.05	0.1	0.2

Theorem 3. *Let $F_{\mathcal{N}}$ denote the repartition function of $\mathcal{N}(0,1)$. The max bias follows the identity:*

$$B_{\omega,\sigma}(\alpha, \Delta T) = 1 - 4\sum_{j\in\mathbb{N}}\left[F_{\mathcal{N}}\left(\frac{j+1-\alpha/2}{\sigma\sqrt{\Delta T}}\right) - F_{\mathcal{N}}\left(\frac{j+\alpha/2}{\sigma\sqrt{\Delta T}}\right)\right]$$

The proof of Theorem 3 is described in Appendix B. It consists of a computation of the probability through the repartition function. Therefore, it corresponds to summing the areas depicted in red and subtracting the area depicted in blue from Figure 4.

Remark 4. The formula described in Theorem 3 highlights that the max bias B is a function of solely $Q = \sigma^2\Delta T$ and α. As a consequence, we will write $B(\alpha, Q)$ to denote $B_{\omega,\sigma}(\alpha, \Delta T)$. The behavior of B depicted in Table 2 is monotonic and follows our expectation:

1. When the volatility Q is fixed, the function $a \to B(a, Q)$ is an increasing function with a minimal value of $B(0, Q)$ and a maximal value of $B(1, Q) = 1$.

2. When the bias a of the duty cycle is fixed, $Q \to B(a, Q)$ is a decreasing function with a maximal value of $B(a, 0) = 1$, and a minimal value of $B(a, \infty) = a$.

Table 3: Max bias estimation $B^*(Q)$ is computed following [BLMT11], and the exact value $B(0.5, Q)$ follows our approach for $\alpha = 0.5$. The bias depicted in red should not be taken into account since the variable is outside the domain of definition; the bias depicted in green shows the points where both estimates coincide.

Q	$B(0.5,Q)$	$B^*(Q)$	Q	$B(0.5,Q)$	$B^*(Q)$
0.005	0.999186096	1.153578017	0.08	0.262481857	0.262482142
0.006	0.997502338	1.131030566	0.09	0.215463331	0.215463379
0.007	0.994385123	1.10892382	0.1	0.17686714	0.176867148
0.008	0.989622785	1.087249165	0.2	0.024568816	0.024568816
0.009	0.983184011	1.065998156	0.3	0.003412882	0.003412882
0.01	0.975161339	1.04516251	0.4	0.000474087	0.000474087
0.015	0.917546335	0.946936107	0.5	$6.5856 \cdot 10^{-5}$	$6.5856 \cdot 10^{-5}$
0.03	0.702200458	0.704257101	0.7	$1.27078 \cdot 10^{-6}$	$1.27078 \cdot 10^{-6}$
0.05	0.47448746	0.474546359	0.9	$2.45213 \cdot 10^{-8}$	$2.45213 \cdot 10^{-8}$
0.06	0.389530294	0.389540261	1	$3.40628 \cdot 10^{-9}$	$3.40628 \cdot 10^{-9}$
0.07	0.319759728	0.319761415			

In comparison, the bias $e(x, \Delta T) = E((-1)^{\mathcal{S}(t_0+\Delta T)}\phi(t_0) = x)|$, for a specific initial phase x both depends on the drift ω and is not monotonic when ΔT increases. Indeed, it fluctuates between local maxima of $B(a, Q)$ when $(x + \omega\Delta T) = \alpha/2 \mod 1$ and local minima of $-B(a, Q)$ when $(x + \omega\Delta T) = (1 + \alpha)/2 \mod 1$ therefore reaching 0 for an intermediate value.

Comparaison with [BLMT11]. As a consequence of this theorem, we derive a process to compute exactly the max bias. Indeed, it suffices to follow the formula and compute the first term of the sum once the precision requirement is met (for instance, we can stop once we have reached $j > 10\sigma\sqrt{\Delta T}$).

Although the work presented in [BLMT11] did not study the maximum bias as an analytical tool, the probability formula provided in that study (Proposition 1) still enables us to derive an approximation for the maximum bias specifically for the case when $\alpha = 1/2$. Indeed:

$$B^*(Q) = \frac{4}{\pi} e^{-2\pi^2 Q}$$

In Table 3, we confront both methods for estimating the max bias. We remark that both functions coincide once Q is sufficiently large ($Q > 0.06$).

Some Properties. We now introduce several bounding applications for the maximum bias within our context, where $\phi_0 \sim \mathbb{U}$. These applications are compiled in Theorem 4.

Theorem 4. *Let's denote by $(t_1, t_2, \cdots, t_k)$ an ordered sequence of sampling time of a ring oscillator whose initial phase follows a uniform distribution. The following inequalities hold:*

1. $\forall\ t_0 \leq t, |\mathbb{E}[(-1)^{\mathcal{S}(t)}]| = a$
2. $\forall\ t \in]t_1, t_2[, |\mathbb{E}[(-1)^{\mathcal{S}(t_1)+\mathcal{S}(t_2)}]| \leq B(\alpha, t_2 - t)B(\alpha, t - t_1)$
3. $\forall\ (t_1, t_2, \cdots, t_k), |\mathbb{E}[(-1)^{\mathcal{S}(t_1)+\cdots+\mathcal{S}(t_k)}]| \leq B(\alpha, t_2 - t_1)B(\alpha, t_k - t_{k-1})$

By studying the extreme cases $t \to t_1$ or $t \to t_2$ in item 2, we obtain the following inequality:

$$\left|\mathbb{E}[(-1)^{\mathcal{S}(t_1)+\mathcal{S}(t_2)}]\right| \leq B(\alpha, t_2 - t_1)$$

The proof of Theorem 4 is given in Appendix C. The proof of the first item is a direct consequence of the definition of α and $\phi_0 \sim \mathbb{U}$. The second and third elements are proven by splitting time and studying the Wiener process with time evolving both in forward and backward directions.

4.3 Probability of a Sampled Sequence

In this section, we introduce a formula along with a practical approximation for determining the probability of a sampled sequence, utilizing the Walsh transform and bias analysis.

More precisely, we examine the probability that the sampled sequence after the n-th measurement, denoted as $(s(i))_{i\leq n}$, corresponds to a specified binary sequence $b_i \in \{0,1\}^n$. Decomposing this probability on its Walsh basis yields the following identity:

$$\mathbb{P}\left[\forall i,\ s(i) = b_i\right] = \frac{1}{2^n}\left(1 + \sum_{w\in\{0,1\}^n\setminus\{0\}} (-1)^{w\cdot b} C(w)\right)$$

where $C(w) = \mathbb{E}\left((-1)^{w_0 s(0)+w_1 s(1)+\cdots+w_n s(n)}\right)$.

This formula can be simplified under any equivalence relation $\mathcal{R}$ that maintains the value of the coefficient $C(w)$. In fact, we can express the probability in the following manner:

$$\mathbb{P}\left[\forall i,\ s(i) = b_i\right] = \frac{1}{2^n}\left(1 + \sum_{I\in\{0,1\}^n\{0\}/\mathcal{R}} \theta_I \sum_{w\in I}(-1)^{w\cdot b}\right),$$

where $\theta_I = C(w)$ for any $w \in I$. Following the result presented in Subsection 4.1, it is clear that the coefficients $C(w)$ and $C(w')$ share the same value if w and w' are connected by some symmetry or translation. Therefore, considering $\mathcal{R}$, the relation generated by the composition of translation $\sim_\tau$ and symmetry $\sim_\mathcal{S}$,

$$\begin{aligned} &w \sim_\tau w' : w = w' \ggg_r \text{ for } r \in \{-n+1, \cdots, n-1\},\\ &w \sim_\mathcal{S} w' : w = Rev(w'),\end{aligned}$$

it is possible to derive the formula presented in Theorem 5.

Theorem 5 (Probability formula derived from max bias analysis)**.** *Given an $(\alpha, \omega, \sigma, \Delta T)-$RO and $(b_i)_{1\leq i\leq n} \in \{0,1\}^n$, the following identity holds:*

$$\mathbb{P}\left[\forall i,\ s(i) = b_i\right] = \frac{1}{2^n}\left(1 + \theta_0 \sum_{i=1}^{n}(-1)^{b_i} + \theta_1 \sum_{i=1}^{n-1}(-1)^{b_i+b_{i+1}} + \delta(\boldsymbol{b})\right)$$

where $\theta_0 = \mathbb{E}\left((-1)^{s(t)}\right) = a$, $|\theta_1| = \left|\mathbb{E}\left((-1)^{s(t)+s(t+\Delta T)}\right)\right| \leq B(\alpha, Q)$, $\boldsymbol{b} = (b_1, \cdots, b_n)$ and $|\delta(\boldsymbol{b})| \leq (2^n - 2n)B(\alpha, Q)^2$.

Proof. This proof consists of studying the equivalent classes of $\mathcal{R}$.

- The first class consists of the elements of Hamming weight 1 and corresponds to the term $\theta_0 \sum_{i=1}^{n}(-1)^{b_i}$. Following the first item presented in Theorem 4, we can state: $\theta_0 = \mathbb{E}\left((-1)^{s(t)}\right) = a$.
- The second one gathers elements of Hamming weight 2 such that the two coordinates that are equal to 1 are consecutive; this corresponds to the term $\theta_1 \sum_{i=1}^{n-1}(-1)^{b_i+b_{i+1}}$. Following the last inequality of Theorem 3, we can state: $|\theta_1| = |\mathbb{E}\left((-1)^{s(t)+s(t+\Delta T)}\right)| \leq B(\alpha, Q)$.

- We now study the remaining $2^n - 2n$ vectors of $\{0,1\}^n \setminus \{0\}$ that do not belong to either of the classes presented above and gather them into $\delta(\mathbf{b})$. For any of these w, we know that there exist two non-consecutive coordinates that are both equal to 1. When the Hamming weight of w is equal to 2, the second item of Theorem 4 allows us to derive $|C(w)| \leq B(\alpha, Q)^2$. When the Hamming weight is at least 3, the third item of Theorem 4 also allows us to derive $|C(w)| \leq B(\alpha, Q)^2$. By gathering all terms together, we obtain $|\delta(\mathbf{b})| \leq (2^n - 2n)B(\alpha, Q)^2$.

□

Corollary 2. *Given $(b_i)_{1\leq i\leq n} \in \{0,1\}^n$, let $e((b_i)_{1\leq i\leq n})$ denote the relative bias associated with $\mathbb{P}[\forall i,\ s(i) = b_i]$. In the special case where $\alpha = 1/2$, we can state:*

$$|e((b_i)_{1\leq i\leq n})| \leq B(\alpha, Q)\left|\sum_{i=1}^{n-1}(-1)^{b_i+b_{i+1}}\right| + (2^n - 2n)B(\alpha, Q)^2.$$

Comparaison with [BLMT11]. Once again, we will compare the result derived through this section with the work done in [BLMT11] that provided a formula for the probability in the special case of $\alpha = 1/2$ and Q large enough. More specifically, we will check if Corollary 2 holds for the formula presented in previous work. Indeed, [BLMT11] (Proposition 1) provides the following formula:

$$\mathbb{P}[\forall i,\ s(i) = b_i] = \frac{1}{2^n}\left(1 + \frac{8}{\pi^2}\left(\sum_{i=1}^{n-1}(-1)^{b_i+b_{i+1}}\right) cos(2\pi\omega\Delta T)e^{-2\pi^2 Q} + O\left(2^n e^{-4\pi^2 Q}\right)\right).$$

Our strategy consists of comparing $B(\alpha, Q)$ with $\frac{8}{\pi^2}e^{-2\pi^2 Q}$ and $(2^n - 2n)B(\alpha, Q)^2$ with $2^n e^{-4\pi^2 Q}$. Let's remind that Table 3 establishes that for Q large enough $B(\alpha, Q) \sim \frac{4}{\pi}e^{-2\pi^2 Q}$. Since $2 < \pi$, it yields:

$$\frac{8}{\pi^2}e^{-2\pi^2 Q} \leq B(\alpha, Q).$$

By squaring this inequality, we obtain the second:

$$e^{-4\pi^2 Q} \leq B(\alpha, Q)^2.$$

As a conclusion, the bound provided in Corollary 2 is compatible with [BLMT11] but Theorem 3 is applicable for other cases, mainly when $\alpha \neq 1/2$ and small Q and provides an upper bound solely depending on α and Q (independent of the drift $\omega\Delta T$).

5 Bounding the Entropy

In this section, we will present the two entropy measures discussed in the recommendations from BSI [KS11, PS22] and NIST [TBK+18]: *Shannon entropy* and *min-entropy*. For both measures, we will use max bias analysis to provide efficient bounds.

5.1 Shannon Entropy

Shannon entropy is a widely recognized concept in information theory and probability, serving as a critical measure of the quality of random number generator (RNG) sources.

Definition 4 (Shannon Entropy)**.** Let's consider an $(\alpha, \omega, \sigma, \Delta T) - RO$ with an associated sampled sequence $(s(n))_{n\in\mathbb{N}^*}$; we call the Shannon entropy of the sequence the quantity:

$$H_n = H\left(s(1), \cdots, s(n)\right).$$

We define the Shannon entropy rate τ_S as the limit:

$$\tau_S = \lim_{n\to\infty} \frac{H_n}{n},$$

together with the associated conditional entropy rate:

$$\tau_S^C = H\left(\mathcal{S}(t+\Delta T)|\phi(t)\right) = \int_x H\left(\mathcal{S}(t+\Delta T)|\phi(t)=x\right) dx.$$

The existence of τ_S comes from the decreasing aspect of the sequence $\frac{H_n}{n}$, note that this property is valid for any stationary process.

5.1.1 A Generic Chain of Bounds

Theorem 6. *Let's consider an $(\alpha, \omega, \sigma, \Delta T) - RO$ with an associated sampled sequence $(s(n))_{n\in\mathbb{N}^*}$, for any $n \in \mathbb{N}$ the following inequality chain holds:*

$$h(B(\alpha, Q)) \leq \tau_S^C \leq \tau_S \leq \frac{H_n}{n} \leq \frac{H_{n-1}}{n-1} \leq \cdots \leq \frac{H_2}{2} \leq H_1 = h(a).$$

where $h(b)$ corresponds to the entropy of a Bernoulli of bias b, hence $h(b) = 1 - \frac{(1+b)\log_2(1+b)}{2} - \frac{(1-b)\log_2(1-b)}{2}$.

The proof of Theorem 6 can be found in Appendix D. It first consists of using the stationary aspect to prove that $\frac{H_n}{n}$ is decreasing (a well-known property of stationary processes also reminded in [Saa21]). Then we prove $\tau_S^C \leq \tau_S$ and finally, we use Proposition 4 to prove $h(B(\alpha, Q)) \leq \tau_S^C$.

Remark 5. When considering Theorem 6 when $Q \to \infty$, we can remark that both bounds $h(B(\alpha, Q))$ and $h(a)$ coincide, hence establishing the tightness of the bound in this context. This corresponds to sampling independent bits with bias a.

5.1.2 A Bound for small n.

While the bound derived in Theorem 6 is applicable for all values of n , we can obtain a more favorable bound when n is small, as demonstrated in the following result.

Proposition 5. *Consider the sequences defined as follow:*

$$\begin{aligned} M_n &= na + (n-1)B(\alpha, Q) + (2^n - 2n)B(\alpha, Q)^2, \\ V_n &= na^2 + (n-1)B(\alpha, Q)^2 + (2^n - 2n)B(\alpha, Q)^4, \\ \Delta_n &= \frac{(1-M_n)\ln(1-M_n) + M_n - M_n^2/2}{\ln 2}. \end{aligned}$$

For n sufficently small such that $M_n \leq 1$,

$$H_n \geq n - \frac{V_n}{2\ln 2} - \Delta_n.$$

The proof of Proposition 5 consists of the application of Theorem 1 to the random variable $(s(1), \cdots, s(n))$.

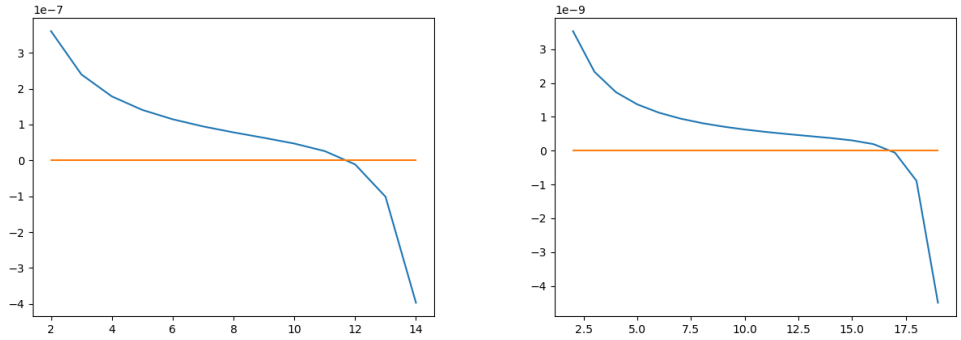

Figure 5: Comparison on both bounds on Shannon entropy: $1 - \frac{V_n}{2n \ln 2} - \Delta_n/n - h(B)$ in function of n for $B = 10^{-3}$ and $a = 10^{-5}$ (resp $B = 10^{-4}$ and $a = 10^{-5}$). Once $n \geq 13$ (resp $n \geq 18$), $h(B)$ provides a better lower bound for entropy.

5.2 Providing Proven Bounds for Min-Entropy

Min-entropy is another crucial measure used in information theory, particularly in the context of assessing randomness and security in random number generators. This measure preferred by the NIST is introduced in the following definition.

Definition 5 (Min-Entropy)**.** Let X be discrete random variable over a finite set, its min-entropy is defined by the following quantity:

$$H_\infty(X) = -\log_2 \left(\max_x \mathbb{P}[X = x] \right).$$

For a given $(\alpha, \omega, \sigma, \Delta T) - RO$, we call min-entropy rate the quantity:

$$\frac{H_\infty(s(1), \cdots, s(n))}{n}$$

and the associated conditional min-entropy rate:

$$\tau_\infty^C = \max_x H_\infty \left(s(1) | \phi(t_0) = x \right).$$

Theorem 7. *Let's consider a ring oscillator with an associated sampled sequence* $(s(n))_{n \in \mathbb{N}^*}$*, for any* $n \in \mathbb{N}$ *the following inequality chain holds:*

$$1 - \log_2(1 + B(\alpha, Q)) \leq \tau_\infty^C \leq \frac{H_\infty(s(1), \cdots, s(n))}{n}.$$

6 Multi-Ring Oscillators

We now study the case of L independent ring oscillator $(\alpha_i, \omega_i, \sigma_i, \Delta T) - (RO_i)_{1 \leq i \leq L}$ whose sampled sequences are processed through a vectorial filtering function $F : \{0,1\}^L \to \{0,1\}^r$, thereby producing the output sequence $(f(n))_{n \in \mathbb{N}} = (F(s_1(n), \cdots, s_L(n)))_{n \in \mathbb{N}}$, with $f(n) \in \{0,1\}^r$.

In this section, we will establish bounds for the Shannon entropy and the min-entropy of the output sequence $(f(n))_{n \in \mathbb{N}}$. Additionally, we will propose efficient filtering functions that optimize the random bitrate generated while adhering to a specified entropy threshold.

6.1 A Vectorial Max Bias

First, we aim at defining a max-bias for the output sequence based on the max bias of each ring oscillator used in the TRNG process. For each ring oscillator, let's denote by ϵ_i as the conditional bias:

$$\epsilon_i(x_i) = \mathbb{E}[(-1)^{s_i(1)}|\phi_i(t_0) = x_i],$$

whose norm is bounded by the associated max bias B_i. For any non-zero vector $\mathbf{u} \in \{0,1\}^r \setminus \{0\}$, we define the conditional bias of the output by:

$$\epsilon_F(\mathbf{u}, \mathbf{x}) = \mathbb{E}[(-1)^{\mathbf{u}\cdot f(1)}|\Phi(t_0) = \mathbf{x}],$$

where $\Phi \in [0,1[^L$ corresponds to the phase vector for all L ring oscillators (hence ϕ_i is the i-th coordinate of Φ). By decomposing on the Walsh Basis, it is possible to rewrite $(-1)^{\mathbf{u}\cdot f(1)}$ as follows:

$$(-1)^{\mathbf{u}\cdot f(1)} = \frac{1}{2^L} \sum_{\mathbf{w}\in\{0,1\}^L} \hat{F}(\mathbf{u}, \mathbf{w})(-1)^{w_1\cdot s_1(1)+\cdots+w_L\cdot s_L(1)},$$

where $\hat{F}$ follows Definition 1. By considering the expected value of this expression and then applying the piling-up lemma and the linearity of the expected value, we derive the following equality:

$$\epsilon_F(\mathbf{u}, \mathbf{x}) = \frac{1}{2^L} \sum_{\mathbf{w}\in\{0,1\}^L} \hat{F}(\mathbf{u}, \mathbf{w}) \prod_{i=1}^{L} \epsilon_i(x_i)^{w_i}.$$

This last expression uses the Walsh transform of the filtering function to connect the conditional bias of each ring oscillator with the conditional bias of the output. From this identity, we will define a notion for the max bias of the output. Indeed, for each ring oscillator, we can bound the norm of the conditional bias by its max bias: $|\epsilon_i| \leq B_i$. As a consequence, the following inequality holds:

$$|\epsilon_F(\mathbf{u}, \mathbf{x})| \leq \frac{1}{2^L} \sum_{\mathbf{w}\in\{0,1\}^L} |\hat{F}(\mathbf{u}, \mathbf{w})| \prod_{i=1}^{L} B_i^{w_i}.$$

Therefore, we define the max bias as the maximal value of $\mathbf{u}$ of the right term:

$$B_F = \max_{\mathbf{u}} \frac{1}{2^L} \sum_{\mathbf{w}\in\{0,1\}^L} |\hat{F}(\mathbf{u}, \mathbf{w})| \prod_{i=1}^{L} B_i^{w_i}.$$

With this definition of max bias, we will provide bounds on the entropy of the output sequence. Although the computation of B_F seems costly for large L, the discussion done in Subsection 6.3 will present some specific and interesting cases of functions F for which this computation simplifies.

6.2 Bounds on Entropies.

We will now present bounds on the Shannon entropy rate and min-entropy rate, based on the maximum bias illustrated in the preceding section. First let's remark that, similarly to the boolean case, the Shannon and the min-entropy rates are bounded by the conditional entropies. More precisely,

$$\frac{H(f(1), \cdots, f(n))}{n} \geq H\left(f(1)|\Phi(t_0)\right) = \tau_S^C,$$
$$\frac{H_\infty(f(1), \cdots, f(n))}{n} \geq \max_{\mathbf{x}\in[0,1[^L} H_\infty\left(f(1)|\Phi(t_0) = \mathbf{x}\right) = \tau_\infty^C.$$

In the following, we will study and provide bounds on these conditional entropies.

6.2.1 Min-Entropy

By utilizing Corollary 1, we can derive a bound for the maximum probability based on the maximum bias. This derivation leads us to Theorem 8.

Theorem 8 (General Lower Bound for Min-Entropy)**.** *If* $B_F \leq \frac{1}{2^r-1}$*, then*

$$\tau_\infty^C \geq r - \log_2\left(1 + (2^r - 1)B_F\right)$$

6.2.2 Shannon Entropy

Our approach regarding the Shannon entropy is built from Theorem 1 and formalized in Theorem 9.

Theorem 9 (Lower Bound for Shannon Entropy of MO-TRNG)**.** *Consider a MO-TRNG* $(\alpha_i, \omega_i, \sigma_i, \Delta T) - (RO_i)_{1\leq i\leq L}$*, we start by defining the family of functions* $(\mathcal{E}_i)_{1\leq i\leq L}$ *as follows:*

$$\mathcal{E}_i : k \mapsto \int_0^1 \epsilon_i(x)^k \, dx,$$

that satisfies $\mathcal{E}_i(0) = 1$*,* $\mathcal{E}_i(1) = a_i$ *and* $\mathcal{E}_i(2) \leq B_i^2$ *. If the max bias of the MO-TRNG satisfies* $B_F \leq \frac{1}{2^r-1}$*, then one can derived the following bounds:*

- ***General Bound.***

$$\tau_S^C \geq r - \frac{C}{2\ln 2} - \Delta((2^r - 1)B_F)$$

where

$$C = \frac{1}{2^{2L}} \sum_{\boldsymbol{u}\in\{0,1\}^r\setminus 0, \boldsymbol{v}\in\{0,1\}^L, \boldsymbol{w}\in\{0,1\}^L} \hat{F}(\boldsymbol{u},\boldsymbol{v})\hat{F}(\boldsymbol{u},\boldsymbol{w}) \prod_{i=1}^{L} \mathcal{E}_i(v_i + w_i).$$

- ***Practical Bounds.***

$$\begin{aligned}\tau_S^C &\geq r - \frac{C'}{2\ln 2} - \Delta((2^r - 1)B_F) \\ &\geq r - \frac{C''}{2\ln 2} - \Delta((2^r - 1)B_F)\end{aligned}$$

where

$$C' = \frac{W_F}{2^{2L}} \prod_{i=1}^{L}(1 + 2a_i + \mathcal{E}_i(2)),$$

$$C'' = \frac{W_F}{2^{2L}} \prod_{i=1}^{L}(1 + 2a_i + B_i^2).$$

with

$$\begin{aligned}W_F &= \max_{\boldsymbol{v}\in\{0,1\}^L} \sum_{\boldsymbol{u}\in\{0,1\}^r\setminus 0} \hat{F}(\boldsymbol{u},\boldsymbol{v})^2 \\ &= \max_{\boldsymbol{v}\in\{0,1\}^L} \left[2^r \sum_{\boldsymbol{a}\in\{0,1\}^r} \left(\sum_{\boldsymbol{x}\in F^{-1}(\boldsymbol{a})} (-1)^{v\cdot x} \right)^2 - 2^{2L}\delta_{\boldsymbol{v}=0} \right]\end{aligned}$$

The proof of this theorem consists of the proofs of both the generic and the practical bounds. First, the proof of the general bound is done by applying Theorem 1 and developing the term $C = \int_{\mathbf{x}\in[0,1[^L} \sum_{\mathbf{u}\neq 0} \epsilon_F(\mathbf{u},\mathbf{x})^2$ to obtain the expression presented in Theorem 9.

Then the proof of the practical bounds can be done by proving $C \leq C' \leq C''$. If $C' \leq C''$ follows $\mathcal{E}_i(2) \leq B_i^2$, the proof of $C \leq C'$ consists of the application of Cauchy-Schwarz to show:

$$\left| \sum_{\mathbf{u}\in\{0,1\}^r\setminus 0,} \hat{F}(\mathbf{u},\mathbf{v})\hat{F}(\mathbf{u},\mathbf{w}) \right| \leq \max_{\mathbf{v}\in\{0,1\}^L} \sum_{\mathbf{u}\in\{0,1\}^r\setminus 0} \hat{F}(\mathbf{u},\mathbf{v})^2$$

Remark 6. For the specific case where the output is restricted to one bit ($r = 1$), there is a better bound:

$$\tau_S^C \geq h(B_F).$$

Corollary 3. *If F is equidistributed then:*

$$\tau_S^C \geq r - \frac{\prod_{i=1}^{L}(1 + 2a_i + \mathcal{E}_i(2))}{2\ln 2} - \Delta((2^r-1)B_F).$$

Indeed, rough estimates on W_F indicate:

$$W_F \leq 2^r \sum_{\mathbf{a}\in\{0,1\}^r} (\sharp F^{-1}(a))^2.$$

Consequently, in the cases where F is equidistributed, $W_F \leq 2^{2L}$, effectively eliminates the impact of F the second term of the bound.

6.3 Choosing the Filtering Function.

In this section, we will discuss choices that can be made for F to simplify the expressions presented before. We will start by studying linear filtering and then bend filtering.

6.3.1 Vectorial Linear Filtering

In the case where F, is linear the following simplifications occur:

$$\mathbf{u}\cdot F(\mathbf{x}) = F^\intercal(\mathbf{u})\cdot\mathbf{x}$$
$$\hat{F}(\mathbf{u},\mathbf{w}) = \delta_{F^\intercal(\mathbf{u})=\mathbf{w}}.$$

As a consequence, the expression of the following parameters is simplified:

$$B_{Linear} = \max_{\mathbf{u}\neq 0,\mathbf{w}=F^\intercal(\mathbf{u})} \prod_{i=1}^{L} B_i^{w_i}$$
$$C_{Linear} = \sum_{\mathbf{u}\neq 0,\mathbf{w}=F^\intercal(\mathbf{u})} \prod_{i=1}^{L} B_i^{2w_i}.$$

Theorem 10 (Lower Bound for Entropies with Linear Filtering)**.** *Consider a MO-TRNG $(\alpha_i,\omega_i,\sigma_i,\Delta T)-(RO_i)_{1\leq i\leq L}$, if $B_{Linear} \leq \frac{1}{2^r-1}$, then*

$$\tau_S^C \geq r - \frac{C_{Linear}}{2\ln 2} - \Delta((2^r-1)B_{Linear}),$$
$$\tau_\infty^C \geq r - \log_2\left(1 + (2^r-1)B_{Linear}\right).$$

In the case where all ring oscillators have a similar order duty cycle α and max bias B_{RO}, then by defining:

$$d = \min_{\mathbf{u}\neq 0, \mathbf{w}=F^{\intercal}(\mathbf{u})} \sum_{i=1}^{L} w_i,$$

the max bias of the output is $B_{Linear} = B_{RO}^d$ and $C_{Linear} \leq (2^r-1)B_{RO}^{2d}$. This allows the simplification presented in the next theorem.

Theorem 11 (Lower Bound for Entropies with Linear Filtering and Similar Ring Oscillators)**.** *Consider a MO-TRNG* $(\alpha_i, \omega_i, \sigma_i, \Delta T) - (RO_i)_{1\leq i\leq L}$*, if* $B_{RO}^d \leq \frac{1}{2^r-1}$*, then*

$$\tau_S^C \geq r - \frac{(2^r-1)B_{RO}^{2d}}{2\ln 2} - \Delta((2^r-1)B_{RO}^d),$$
$$\tau_\infty^C \geq r - \log_2\left(1 + (2^r-1)B_{RO}^d\right).$$

This theorem illustrates how it is possible to achieve any target entropy by increasing d.

Remark 7. We can remark that choosing F corresponds to the choice of a code with minimal distance d. To achieve the optimal entropy bound, it is crucial to minimize the associated bias. Therefore, we aim to identify the linear code with the greatest minimal distance, which we determined using the `BestKnownLinearCode` function from the Magma library.

6.3.2 Vectorial Bent Filtering

Vectorial Bent functions have the complete opposite behavior of linear functions and solely exist when $L = 2n$ and $r \leq n$ (as proven in [Nyb91]). Indeed, the Walsh spectrum of any of its components is uniformly spread on all the 2^L values instead of being reduced to one point in the case of a linear map. A simple example of such a vectorial bent function is the quadratic function:

$$\begin{array}{rccc} F: & \mathbb{F}_{2^{2n}} & \to & \mathbb{F}_{2^n} \\ & (x_1, x_2) & \mapsto & x_1 x_2 \end{array}$$

As a consequence of the Walsh behavior, the following simplifications occur:

$$|\hat{F}(\mathbf{u}, \mathbf{w})| = 2^{L/2}.$$

Hence, the expression of the following parameters is simplified:

$$B_{Bent} = \frac{1}{\sqrt{2^L}} \prod_{i=1}^{L} (1 + B_i)$$
$$C_{Bent} = \frac{2^r - 1}{2^L} \prod_{i=1}^{L} (1 + a_i + B_i^2).$$

Theorem 12 (Lower Bound for Entropies with Bent Filtering)**.** *Consider a MO-TRNG* $(\alpha_i, \omega_i, \sigma_i, \Delta T) - (RO_i)_{1\leq i\leq L}$*, if* $B_{Bent} \leq \frac{1}{2^r-1}$*, then*

$$\tau_S^C \geq r - \frac{C_{Bent}}{2\ln 2} - \Delta((2^r-1)B_{Bent}),$$
$$\tau_\infty^C \geq r - \log_2\left(1 + (2^r-1)B_{Bent}\right).$$

If we can assume that all ring oscillators have a similar order duty cycle α and max bias B_{RO}, the max bias of the output is $B_{Bent} = \frac{(1+B_{RO})^L}{\sqrt{2^L}}$ and $C_{Bent} \leq (2^r-1)\left(\frac{1+2a+B_{RO}}{2}\right)^L$. This allows the simplification presented in the next theorem.

Theorem 13 (Lower Bound for Entropies with Bent Filtering and Similar Ring Oscillators)**.** *Consider a MO-TRNG* $(\alpha_i, \omega_i, \sigma_i, \Delta T) - (RO_i)_{1\leq i\leq L}$*, if* $\frac{(1+B_{RO})^L}{\sqrt{2^L}} \leq \frac{1}{2^r-1}$*, then*

$$\tau_S^C \geq r - \frac{(2^r-1)}{2\ln 2}\left(\frac{1+2a+B_{RO}}{2}\right)^L - \Delta((2^r-1)\frac{(1+B_{RO})^L}{\sqrt{2^L}}),$$
$$\tau_\infty^C \geq r - \log_2\left(1+(2^r-1)\frac{(1+B_{RO})^L}{\sqrt{2^L}}\right).$$

One can remark that this choice of filtering function may fail to meet specific Shannon entropy targets when the relative duty cycle a is high. For example, if $a > 1/3$, then $\frac{1+2a+B_{RO}}{2} > 1$, leading to suboptimal bounds. In the subsequent applications, we will focus on linear filtering, as it offers the best bounds.

Remark 8. Note that the established entropy bound is unchanged if we use $F \circ G$ rather than F when G is invertible.

6.4 Applications

In this section, we will leverage the tools discussed in this work to present results on bitrate optimization within the context of linear filtering, which has demonstrated the most promising results. To simplify the proposed applications, we will consider a MO-TRNG $(\alpha_i, \omega_i, \sigma_i, \Delta T) - (RO_i)_{1\leq i\leq L}$ where all ring oscillators have similar parameters, i.e

$$\alpha \simeq \alpha_1 \simeq \cdots \simeq \alpha_L$$
$$\sigma \simeq \sigma_1 \simeq \cdots \simeq \sigma_L.$$

For a given number of ring oscillators L, our objective is to evaluate all potential dimensions for the output and exhibit the best choice, i.e. to explore the value of r. As outlined in Subsubsection 6.3.1, for a given r (and L) we select the associated linear code to maximize the minimum distance since they provide the best entropy bounds. The target entropy follows the recommendation of the BSI [PS22] that stipulates that the entropy per bit generated should either have a Shannon entropy of at least 0.998 or a min-entropy of at least 0.98. For each choice of r, max bias analysis allows to determine a quality factor (thus a sampling period) $Q_r = \sigma\Delta T_r$ that meets the prescribed entropy constraints. Consequently, by comparing this sampling period to that of the boolean XOR operation, we can derive the time ratio $\tau_r = \Delta T_r/\Delta T_1$. The output bitrate is then calculated as the ratio r/τ_r.

In Figure 6, Figure 7, and Figure 8, we visualize the bitrate as a function of r and α, with the targeted min-entropy of 0.98 represented on the left and the Shannon entropy of 0.998 on the right. Notably, it appears that α has minimal impact on the value of r that maximizes the bitrate. Therefore, for a specified number of ring oscillators, an optimal size for the output sequence can be associated.

7 Conclusion

Max analysis has proven to be the optimal level of abstraction for the MO-TRNG construction, as it yields efficient, verified bounds and facilitates vectorial post-processing that guarantees both the efficiency and resilience of the random output. For practical applications, this technique is straightforward to implement, as calculating entropy can be accomplished using basic resources such as Excel, thus facilitating the work of hardware designers and evaluators.

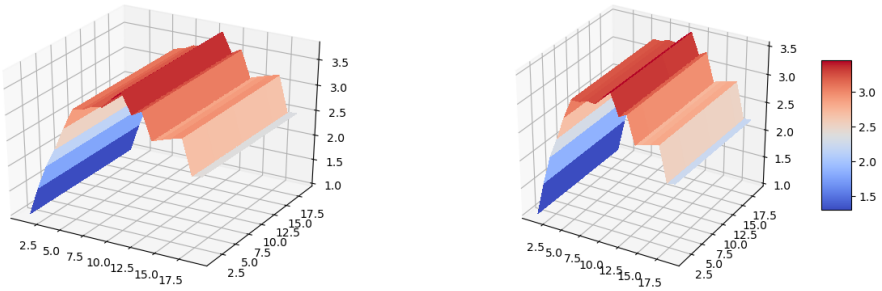

Figure 6: Output bitrate with 32 RO with targeted min-entropy equal to 0.98 (resp Shannon entropy equal to 0.998) in function of a and r

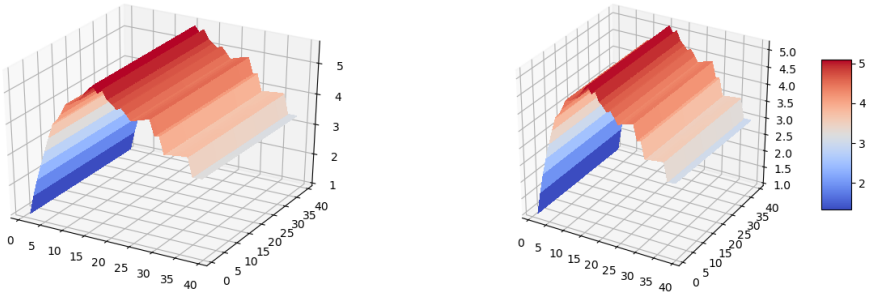

Figure 7: Output bitrate with 64 RO with targeted min-entropy equal to 0.98 (resp Shannon entropy equal to 0.998) in function of a and r

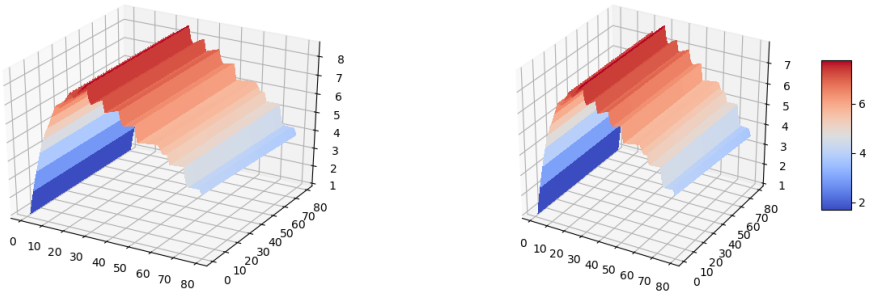

Figure 8: Output bitrate with 128 RO with targeted min-entropy equal to 0.98 (resp Shannon entropy equal to 0.998) in function of a and r

Future Work. There may be opportunities to explore the application of max bias analysis in various other contexts. For example, we could investigate its use in models that incorporate flicker noise or even in entirely different sources of randomness. For instance, one could apply this analysis to derive entropy bounds for constructions with independent sources of randomness modeled with a hidden Markov process or hidden Wiener process.

Another avenue for exploration involves studying post-processing methods with different objectives. For instance, it would be interesting to consider the ideal vectorial post-processing techniques that optimize bitrate in the context of potential malfunctions, as such issues may arise in real-world applications.

Acknowledgments. We would like to express our sincere gratitude to David Lubicz for introducing our team to Ring Oscillator-based TRNG, specifically his insights on the Wiener modeling of thermal noise. We extend our thanks to Alexandre Stevanovic for his help in the graphical representation of the applications.

References

[BCF+24] Licinius Benea, Mikael Carmona, Viktor Fischer, Florian Pebay-Peyroula, and Romain Wacquez. Impact of the flicker noise on the ring oscillator-based trngs. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2024(2):870–889, 2024.

[BLMT11] Mathieu Baudet, David Lubicz, Julien Micolod, and André Tassiaux. On the security of oscillator-based random number generators. *Journal of Cryptology*, 24(2):398–425, April 2011.

[HTBF14] Patrick Haddad, Yannick Teglia, Florent Bernard, and Viktor Fischer. On the assumption of mutual independence of jitter realizations in p-trng stochastic models. In *2014 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pages 1–6. IEEE, 2014.

[KG04] Paul Kohlbrenner and Kris Gaj. An embedded true random number generator for fpgas. In *Proceedings of the 2004 ACM/SIGDA 12th international symposium on Field programmable gate arrays*, pages 71–78, 2004.

[KS08] Wolfgang Killmann and Werner Schindler. A design for a physical RNG with robust entropy estimators. In Elisabeth Oswald and Pankaj Rohatgi, editors, *CHES 2008*, volume 5154 of *LNCS*, pages 146–163. Springer, Berlin, Heidelberg, August 2008.

[KS11] Wolfgang Killmann and Werner Schindler. A proposal for: Functionality classes for random number generators. *ser. BDI, Bonn*, 2011.

[LF24] David Lubicz and Viktor Fischer. Entropy computation for oscillator-based physical random number generators. *Journal of Cryptology*, 37(2):13, April 2024.

[Mat93] Mitsuru Matsui. Linear cryptanalysis method for des cipher. In *Workshop on the Theory and Application of of Cryptographic Techniques*, pages 386–397. Springer, 1993.

[Nyb91] Kaisa Nyberg. Perfect nonlinear s-boxes. In *Workshop on the Theory and Application of of Cryptographic Techniques*, pages 378–386. Springer, 1991.

[PS22] Matthias Peter and Werner Schindler. A proposal for functionality classes for random number generators version 2.35 [draft]. *ser. BDI, Bonn*, 2022.

[PV24] Adriaan Peetermans and Ingrid Verbauwhede. Trng entropy model in the presence of flicker fm noise. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2024(4):285–306, 2024.

[Saa21] Markku-Juhani O Saarinen. On entropy and bit patterns of ring oscillator jitter. In *2021 Asian Hardware Oriented Security and Trust Symposium (AsianHOST)*, pages 1–6. IEEE, 2021.

[Sko24] Maciej Skorski. Modelling 1/f noise in trngs via fractional brownian motion. *arXiv preprint arXiv:2410.14205*, 2024.

[Ste04] Vladimir Andreevitch Steklov. *Sur certaines égalités générales communes à plusieurs séries de fonctions souvent employées dans l'analyse, par W. Stekloff...* J. Glasounof, 1904.

[TBK+18] Meltem Sönmez Turan, Elaine Barker, John Kelsey, Kerry A McKay, Mary L Baish, Mike Boyle, et al. Recommendation for the entropy sources used for random bit generation. *NIST Special Publication*, 800(90B):102, 2018.

[VD10] Michal Varchola and Milos Drutarovsky. New high entropy element for fpga based true random number generators. In *Cryptographic Hardware and Embedded Systems, CHES 2010: 12th International Workshop, Santa Barbara, USA, August 17-20, 2010. Proceedings 12*, pages 351–365. Springer, 2010.

A Proof of Theorem 1

Proof. For this proof, we will introduce the moment of order k of e:

$$M_k(e) = \frac{1}{2^r} \sum_x e(x)^k.$$

Going back to the definition of entropy, we can state:

$$H(X) = \frac{1}{2^r} \sum_{x \in \{0,1\}^r} -\log_2(\frac{1+e(x)}{2^r}) \cdot (1+e(x))$$

If we interpret now the log part as a power series (note that we are within the radius of convergence since $B \leq 1$), and after some elementary computations, we can rewrite the entropy as follows :

$$H(X) = r + (r + \frac{1}{\ln 2}) M_1(e) - \frac{1}{\ln 2} \sum_{k \geq 2} \frac{1}{(k-1)k} (-1)^k M_k(e).$$

Note that the $\ln 2$ terms appear since the log we are considering is in base 2. Looking at this new expression of $H(X)$, we can simplify it by reminding that $M_1(e) = 0$ since $\sum_x p(x) = 1 \implies \sum_x e(x) = 0$. Also since we suppose that $|e(x)| \leq B$ for all x, the following inequalities hold:

$$\begin{aligned} H(X) &= r - \frac{M_2(e)}{2\ln 2} - \frac{1}{\ln 2} \sum_{k \geq 3} \frac{1}{(k-1)k} (-1)^k M_k(e) \\ &\geq r - \frac{M_2(e)}{2\ln 2} - \frac{1}{\ln 2} \sum_{k \geq 3} \frac{1}{(k-1)k} B^k \\ &= r - \frac{M_2(e)}{2\ln 2} - \Delta(B) \end{aligned}$$

□

B Proof of Theorem 3

Proof. It is possible to write $B(\Delta T)$ as follows:

$$\begin{aligned}
B(\Delta T) &= \left|\mathbb{E}[(-1)^{\mathcal{S}(t_0+\Delta T)}|\phi(t_0)=x^*]\right| \\
&= |\mathbb{P}[\mathcal{S}(t_0+\Delta T)=0|\phi(t_0)=x^*] - \mathbb{P}[\mathcal{S}(t_0+\Delta T)=1|\phi(t_0)=x^*]| \\
&= |\mathbb{P}[\phi(t_0+\Delta T) \mod 1 \in [0,\alpha[|\phi(t_0)=x^*] - \mathbb{P}[\phi(t_0+\Delta T) \mod 1 \in [\alpha,1[|\phi(t_0)=x^*]|
\end{aligned}$$

Since, $\phi(t_0+\Delta T)-\phi(t_0) \sim \mathcal{N}(\omega\Delta T, \sigma^2\Delta T)$, we can derive an expression for the probability:

$$\begin{aligned}
&\mathbb{P}[\phi(t_0+\Delta T) \mod 1 \in [0,\alpha[|\phi(t_0)=x^*] \\
&= \mathbb{P}[\phi(t_0+\Delta T)-\phi(t_0) \mod 1 \in [\omega\Delta T-\alpha/2, \omega\Delta T+\alpha/2[|\phi(t_0)=x^*] \\
&= \mathbb{P}_{X\leftarrow\mathcal{N}(0,\sigma^2\Delta T)}[X \mod 1 \in [-\alpha/2,\alpha/2[\,] \\
&= \mathbb{P}_{X\leftarrow\mathcal{N}(0,\sigma^2\Delta T)}[X \mod 1 \in [-\alpha/2,0[\cup[0,\alpha/2[\,] \\
&= \sum_{j\in\mathbb{Z}} F_{\mathcal{N}}\left(\frac{\alpha/2+j}{\sigma^2\Delta T}\right) - F_{\mathcal{N}}\left(\frac{j}{\sigma^2\Delta T}\right) + F_{\mathcal{N}}\left(\frac{1+j}{\sigma^2\Delta T}\right) - F_{\mathcal{N}}\left(\frac{1-\alpha/2+j}{\sigma^2\Delta T}\right) \\
&= 2\sum_{j\in\mathbb{N}} F_{\mathcal{N}}\left(\frac{\alpha/2+j}{\sigma^2\Delta T}\right) - F_{\mathcal{N}}\left(\frac{j}{\sigma^2\Delta T}\right) + F_{\mathcal{N}}\left(\frac{1+j}{\sigma^2\Delta T}\right) - F_{\mathcal{N}}\left(\frac{1-\alpha/2+j}{\sigma^2\Delta T}\right)
\end{aligned}$$

Similarly, one can prove:

$$\mathbb{P}[\phi(t_0+\Delta T) \mod 1 \in [\alpha,1[|\phi(t_0)=x^*] = 2\sum_{j\in\mathbb{N}} F_{\mathcal{N}}\left(\frac{1-\alpha/2+j}{\sigma^2\Delta T}\right) - F_{\mathcal{N}}\left(\frac{\alpha/2+j}{\sigma^2\Delta T}\right)$$

Hence, we obtain:

$$\begin{aligned}
B(\Delta T) = 2\sum_{j\in\mathbb{N}} &F_{\mathcal{N}}\left(\frac{\alpha/2+j}{\sigma^2\Delta T}\right) - F_{\mathcal{N}}\left(\frac{j}{\sigma^2\Delta T}\right) + F_{\mathcal{N}}\left(\frac{1+j}{\sigma^2\Delta T}\right) \\
&- F_{\mathcal{N}}\left(\frac{1-\alpha/2+j}{\sigma^2\Delta T}\right) - F_{\mathcal{N}}\left(\frac{1-\alpha/2+j}{\sigma^2\Delta T}\right) + F_{\mathcal{N}}\left(\frac{\alpha/2+j}{\sigma^2\Delta T}\right),
\end{aligned}$$

that reduces to:

$$\begin{aligned}
B(\Delta T) &= -4\sum_{j\in\mathbb{N}}\left(F_{\mathcal{N}}\left(2\frac{1-\alpha/2+j}{\sigma^2\Delta T}\right) - F_{\mathcal{N}}\left(\frac{\alpha/2+j}{\sigma^2\Delta T}\right)\right) + 2\sum_{j\in\mathbb{N}}\left(F_{\mathcal{N}}\left(\frac{j+1}{\sigma^2\Delta T}\right) - F_{\mathcal{N}}\left(\frac{j}{\sigma^2\Delta T}\right)\right) \\
&= 1 - 4\sum_{j\in\mathbb{N}}\left(F_{\mathcal{N}}\left(2\frac{1-\alpha/2+j}{\sigma^2\Delta T}\right) - F_{\mathcal{N}}\left(\frac{\alpha/2+j}{\sigma^2\Delta T}\right)\right)
\end{aligned}$$

□

C Proof of Theorem 4

Proof. The proof of this lemma consists in the proof of all three claims.

1. Let's remind that at any time $t \geq t_0$, the phase at time t, $\phi(t)$ follows a uniform distribution. Hence, we obtain:

$$\begin{aligned}
|\mathbb{E}[(-1)^{\mathcal{S}(t)}]| &= |\mathbb{P}[\mathcal{S}(t)=0] - \mathbb{P}[\mathcal{S}(t)=1]| \\
&= |2\alpha - 1|
\end{aligned}$$

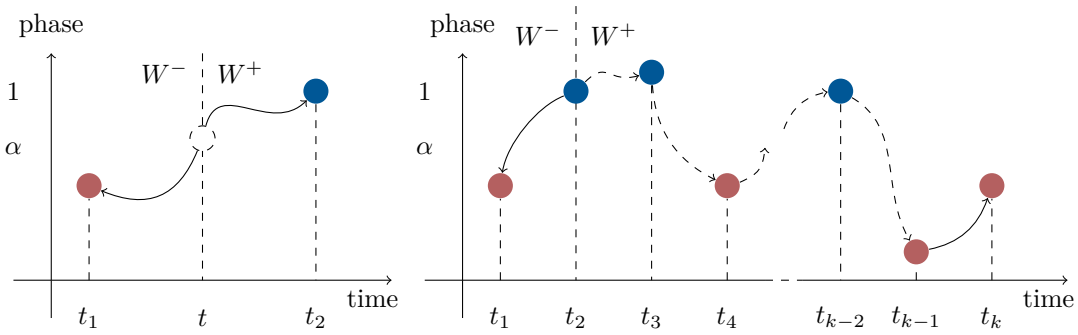

Figure 9: Phase approach with W^+ and W^- in the context of Theorem 4.

2. This proof consists in studying the evolution of the phase while considering the uniform phase at time $t \in]t_1, t_2[$ as the starting point. Therefore, we study two independent paths produced by the Wiener process: $t \to t_1$ and $t \to t_2$. This idea is depicted in the leftmost figure of Figure 9. More precisely, one can write :

$$\begin{aligned}
\left|\mathbb{E}[(-1)^{\mathcal{S}(t_1)+\mathcal{S}(t_2)}]\right| &= \left|\int_{x\in[0,1]} \mathbb{E}[(-1)^{\mathcal{S}(t_1)+\mathcal{S}(t_2)}|\phi(t)=x]\cdot\mathbb{P}[\phi(t)=x]\right| \\
&= \left|\int_{x\in[0,1]} \mathbb{E}[(-1)^{\mathcal{S}(t_1)}|\phi(t)=x]\cdot\mathbb{E}[(-1)^{\mathcal{S}(t_2)}|\phi(t)=x]\cdot\mathbb{P}[\phi(t)=x]\right| \\
&\leq \int_{x\in[0,1]} B(t_2-t)B(t-t_1)\cdot\mathbb{P}[\phi(t)=x] \\
&= B(t_2-t)B(t-t_1)
\end{aligned}$$

3. Similarly to the proof of previous item, we will study the signals with two Wiener executions. As depicted in the rightmost figure of Figure 9, the initial time is placed at t_2 and the considered paths are $t_2 \to t_1$ and $t_2 \to t_k$. This decomposition allows to write:

$$\begin{aligned}
&\left|\mathbb{E}[(-1)^{\mathcal{S}(t_1)+\cdots+\mathcal{S}(t_k)}]\right| \\
&= \left|\int_{x\in[0,1]} \mathbb{E}[(-1)^{\mathcal{S}(t_1)+\cdots+\mathcal{S}(t_k)}|\phi(t_2)=x]\cdot\mathbb{P}[\phi(t_2)=x]\right| \\
&= \left|\int_{x\in[0,1]} \mathbb{E}[(-1)^{\mathcal{S}(t_1)}|\phi(t_2)=x]\cdot(-1)^{\mathcal{S}(t_2)}\cdot\mathbb{E}[(-1)^{\mathcal{S}(t_3)+\cdots+\mathcal{S}(t_k)}|\phi(t_2)=x]\cdot\mathbb{P}[\phi(t_2)=x]\right| \\
&\leq \int_{x\in[0,1]} B(t_2-t_1)\cdot\left|\mathbb{E}[(-1)^{\mathcal{S}(t_3)+\cdots+\mathcal{S}(t_k)}|\phi(t_2)=x]\right|\cdot\mathbb{P}[\phi(t_2)=x]
\end{aligned}$$

The next step consist in decomposing the term $\left|\mathbb{E}[(-1)^{\mathcal{S}(t_3)+\cdots+\mathcal{S}(t_k)}|\phi(t_2)=x]\right|$ alongside the random variable $(\phi(t_3),\cdots,\phi(t_{k-1})|\phi(t_2)=x)$. Thus we can rewrite:

$$\begin{aligned}
&\left|\mathbb{E}[(-1)^{\mathcal{S}(t_3)+\cdots+\mathcal{S}(t_k)}|\phi(t_2)=x]\right| \\
&\leq \int \left|\mathbb{E}[(-1)^{\mathcal{S}(t_k)}|\bigwedge_{2\leq i\leq k-1}\phi(t_i)=x_i]\mathbb{P}[(\phi(t_3)=x_3\wedge\cdots,\phi(t_{k-1})=x_{k-1}|\phi(t_2)=x)]\right|
\end{aligned}$$

On the other hand, the definition of the phase using the Wiener process implies that regarding the phase $\phi(t_k)$, the sole knowledge of the phase $\phi(t_k)$ yields as much information as the knowledge of all the phases $\phi(t_3), \cdots, \phi(t_{k-1})$. Hence,

$$\begin{aligned}
&\left|\mathbb{E}[(-1)^{\mathcal{S}(t_3)+\cdots+\mathcal{S}(t_k)}|\phi(t_2)=x]\right| \\
\leq &\int \left|\mathbb{E}[(-1)^{\mathcal{S}(t_k)}|\bigwedge_{2\leq i\leq k-1}\phi(t_i)=x_i]\mathbb{P}[(\phi(t_3)=x_3\wedge\cdots,\phi(t_{k-1})=x_{k-1}|\phi(t_2)=x)]\right| \\
\leq &\int \left|\mathbb{E}[(-1)^{\mathcal{S}(t_k)}|\phi(t_{k-1})=x_{k-1}]\mathbb{P}[(\phi(t_3)=x_3\wedge\cdots,\phi(t_{k-1})=x_{k-1}|\phi(t_2)=x)]\right| \\
\leq &\int |B(t_k-t_{k-1})\mathbb{P}[(\phi(t_3)=x_3\wedge\cdots,\phi(t_{k-1})=x_{k-1}|\phi(t_2)=x)]| \\
= &B(t_k-t_{k-1}).
\end{aligned}$$

Thus, by incorporating this intermediate result, we arrive at the following inequality:

$$\begin{aligned}
\left|\mathbb{E}[(-1)^{\mathcal{S}(t_1)+\cdots+\mathcal{S}(t_k)}]\right| &\leq \int_{x\in[0,1]} B(t_2-t_1)\cdot\left|\mathbb{E}[(-1)^{\mathcal{S}(t_3)+\cdots+\mathcal{S}(t_k)}|\phi(t_2)=x]\right|\cdot\mathbb{P}[\phi(t_2)=x] \\
&\leq \int_{x\in[0,1]} B(t_2-t_1)B(t_k-t_{k-1})\cdot\mathbb{P}[\phi(t_2)=x] \\
&= B(t_2-t_1)B(t_k-t_{k-1})
\end{aligned}$$

□

Remark 9. The proof presented for item 3 could be done for any starting point in $[t_2, t_{k-1}]$, yet it would still provide the same result in the end.

D Proof of Theorem 6

Proof. We will present the proof of this theorem in three parts. First we will prove that $\frac{H_n}{n}$ is decreasing, then $\tau_S^C \leq \tau_S$ and finally $h(B(Q)) \leq \tau_S^C$.

$\frac{H_n}{n}$ is decreasing. This statement is well known, we will present a proof for self completeness. Let's consider $n-1$ independent ring oscillators each producing a sampling sequence $(s_i(j))1 \leq i \leq n-1, j \in \mathbb{N}$. Since the sequence are independent, we can state:

$$H\left[(s_i(1),\cdots,s_i(n))_{1\leq i\leq n-1}\right] = (n-1)H_n$$

By considering the n-th sampled bits alone, it is possible to write:

$$\begin{aligned}
H\left[(s_i(1),\cdots,s_i(n))\right] &= H\left[(s_i(1),\cdots,s_i(n-1))\right] + H\left[s_i(n)|(s_i(1),\cdots,s_i(n-1))\right] \\
&= H_{n-1} + H\left[s_i(n)|(s_i(1),\cdots,s_i(n-1))\right]
\end{aligned}$$

The idea now is to apply the chain rule to the term $H\left[s_i(n)|(s_i(1),\cdots,s_i(n-1))\right]$ to build H_{n-1}. By remarking

$$\begin{aligned}
H\left[s_i(n)|(s_i(1),\cdots,s_i(n-1))\right] &\leq H\left[s_i(n)|(s_i(n-i),\cdots,s_i(n-1))\right] \\
&= H\left[s_n(i)|(s_n(1),\cdots,s_n(i-1))\right]
\end{aligned}$$

by summing over all $i \in \{1,\cdots,n-1\}$, we obtain:

$$\begin{aligned}
\sum_i H\left[s_i(n)|(s_i(1),\cdots,s_i(n-1))\right] &\leq \sum_i H\left[s_n(i)|(s_n(1),\cdots,s_n(i-1))\right] \\
&= H_{n-1}
\end{aligned}$$

As a consequence:

$$\begin{aligned}(n-1)H_n &= (n-1)H_{n-1} + \sum_i H\left[s_i(n)|(s_i(1),\cdots,s_i(n-1))\right] \\ &\leq (n-1)H_{n-1} + H_{n-1} \\ &= nH_{n-1}.\end{aligned}$$

$\boldsymbol{\tau_S^C \leq \tau_S}$**.** Let's follow the inequalities:

$$\begin{aligned}H_n &= H\left(s(1),\cdots,s(n)\right) \\ &= H\left(s(1)\right) + H(s(2)|s(1)) + \cdots + H(s(n)|s(n-1),\cdots,s(1)) \\ &\geq H\left(s(1)|\phi(t_0)\right) + H(s(2)|s(1),\phi(t_0+\Delta T)) + \cdots + H(s(n)|s(n-1),\cdots,s(1),\phi(t_0+(n-1)\Delta T) \\ &= H\left(s(1)|\phi(t_0)\right) + H(s(2)|\phi(t_0+\Delta T)) + \cdots + H(s(n)|\phi(t_0+(n-1)\Delta T)) \\ &= nH\left(s(1)|\phi(t_0)\right) \\ &= n\tau_S^C\end{aligned}$$

By dividing both term by n and considering $n \to \infty$, we obtain $\tau_S^C \leq \tau_S$.

$\boldsymbol{h(B(Q)) \leq \tau_S^C}$**.** This proof consists in applying Proposition 4 to the random variable $\mathcal{S}(t+\Delta T)|\phi(t) = x$ for all x. Since we know

$$|\mathbb{E}((-1)^{s(t+\Delta T)}|\phi(t) = x)| \leq B(Q)$$

As a consequence of Proposition 4,

$$\begin{aligned}\tau_S^C &= \int_x H((s(t+\Delta T)|\phi(t) = x)dx \\ &\geq h(B(Q)).\end{aligned}$$

□

E Proof of Theorem 7

Proof. By definition of the max Bias:

$$\forall s, x: \quad \mathbb{P}[s(1) = s|\phi(t_0) = x] \leq \frac{1}{2}(1 + B(\alpha, Q)).$$

Therefore

$$H_\infty\left(s(1)|\phi(t_0) = x\right) \geq 1 - \log_2(1 + B(\alpha, Q)).$$

Considering this inequality for $\max \phi$, we obtain $1 - \log_2(1 + B(\alpha, Q)) \leq \tau_\infty^C$.

Were are then left to prove that

$$\tau_\infty^C \leq \frac{H_\infty(s(1),\cdots s(n))}{n}$$

by defining $\mathcal{D}[b_1,\cdots b_i]$ the distribution on the phase at time t_i: $\phi(t_i)$ that satisfies the following:

$$\mathbb{P}\left[s(i) = b_i|s_1 = b_1,\cdots,s_{i-1} = b_{i-1}\right] = \mathbb{P}\left[s(i) = b_i|\phi(t_{i-1}) \leftarrow \mathcal{D}[b_1,\cdots b_{i-1}]\right],$$

it is possible to derive the following equations:

$$\begin{aligned}\mathbb{P}\left[s_1 = b_1, \cdots, s_i = b_i\right] &= \prod_i \mathbb{P}\left[s_i = b_i | s_1 = b_1 \cdots, s_{i-1} = b_{i-1}\right] \\ &= \prod_i \mathbb{P}\left[s_i = b_i | \phi(t_i) \leftarrow \mathcal{D}[b_1, \cdots b_{i-1}]\right] \\ &\leq (2^{-\tau_\infty^C})^n\end{aligned}$$

thereby proving

$$\tau_\infty^C \leq \frac{H_\infty(s(1), \cdots s(n))}{n}.$$

□

F Proof of Theorem 8

Proof. Our analysis consists in invoking Corollary 1 to asses :

$$\epsilon_F(\mathbf{u}, \mathbf{x}) \leq B_F \implies e_F(\mathbf{v}, \mathbf{x}) \leq (2^r - 1)B_F$$

Since

$$\mathbb{P}(f(1) = \mathbf{v} | \Phi(t_0) = \mathbf{x}) = \frac{1}{2^r}(1 + e_F(\mathbf{v}, \mathbf{x}))$$

then

$$\tau_\infty^C = \max_{\mathbf{x} \in [0,1[^L} H_\infty\left(f(1) | \Phi(t_0) = \mathbf{x}\right) \geq r - \log_2\left(1 + (2^r - 1)B_F\right).$$

□