Binning, Generalized von Neumann and XOR, von Neumann Procedure — Digitization and mathematical post-processing in (Q)RNGs

Torsten Schütze
Rohde & Schwarz SIT GmbH
Stuttgart/Germany

DGA Workshop “Random number generators and PUF”
European Cyber Week, Rennes, France
November 21, 2024

ROHDE&SCHWARZ
Make ideas real

Overview

(i) QRNG workshop I, BSI, Bonn, 12/2018: *Experiences with the evaluation of PTRNGs*

- Overall evaluation of a Zener diode based RNG as class PTG.3
- PTRNG: Physical True Random Number Generator, Quantum RNGs (QRNGs) are a subset

(ii) QRNG workshop II, Fraunhofer IOF, Jena, 01/2020: *Some thoughts about post-processing in TRNGs*

- Overview of mathematical post-processing and experiences with a (too heavy) (cryptographic) post-processing; unbiasing methods for *independent* bits
- Rich theory, many methods, but all / most for *independent and identically distributed (i.i.d.) bits, that are biased*; *no dependency*

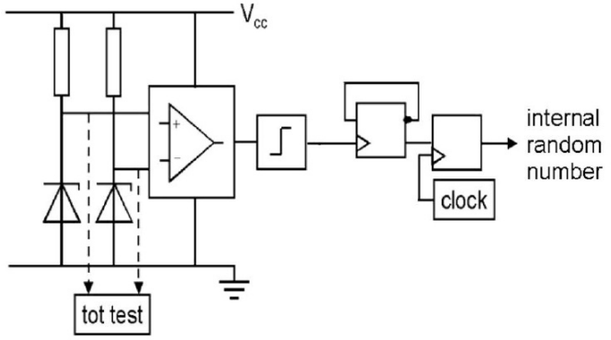
(iii) 806. WE-Heraeus-Seminar on Physics and Security – from Random Numbers to Secure Communication, Bad Honnef, 03/2024 and this talk Rennes, 11/2024: *Binning, Generalized von Neumann and XOR, von Neumann Procedure — Digitization and mathematical post-processing in (Q)RNGs*

- experiences with well-known post-processing methods in case of *dependencies* and *perturbations*
- all results from practical experiences in RNG evaluation, when things aren't going so well

Running examples for illustration

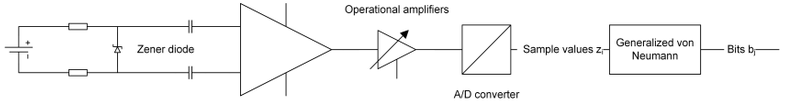
TRNG by F. Bergmann, Berlin

- Two noisy (matched pairs) Zener diodes in differential mode
- Discrete random signal = number of 0-1-crossings in Schmitt trigger
- Stochastic model $\Longrightarrow$ W. Killmann, W. Schindler: *A design for a physical RNG with a robust entropy estimator*, CHES 2008.



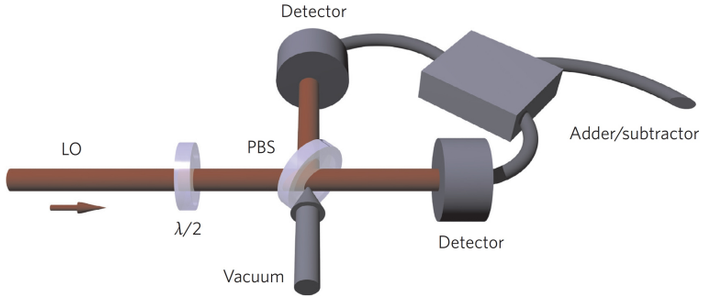
TRNG by Rohde & Schwarz SIT

- One noisy Zener diode (avalanche noise)
- Discrete random signal = digitized sample values z_i after A/D converter
- Random raw bits b_j after Generalized von Neumann
- Approved PTG.3 for harsh environmental conditions



QRNG by Max Planck Institute for the Science of Light

- Homodyne detection of lowest energy vacuum state
- C. Gabriel et al.: *A generator for unique quantum random numbers based on vacuum states*, Nature Photonics 2010.197.



Why these RNGs? What do they have in common?

- I know them well. Of course, I know RNG 2 best.
- QRNG 3 is actually used: BMBF project “Chip-basiertes Quantenzufalls Device – CBQD” and KeeQuant / OHB
- They are all quite similar in a certain sense:
- RNG 1 discretizes the analogue random signal (difference signal of avalanche noise of Zener diodes) in time direction; number of 0-1-crossings in Schmitt trigger
  - important: realizations of a *q-dependent stationary process*
  - probability density distribution of times between consecutive 0-1-crossings $\approx$ Gamma distribution with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$

$$f(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) \quad \text{for } x > 0$$

- RNG 2 discretizes the analoque random signal (avalanche noise of one Zener diode) at equidistant points with k-bit ADC to get sample values z_i; amplitude direction
  - random bits b_j are raw bits after Generalized von Neumann procedure
  - important: stationary process (time-local stationarity), difference of sample values is normally distributed (in the limit case $k \to \infty$)
  - stochastic model for bits: bits are realizations of a Bernoulli process with one-step dependency, parameters bias $p - 0.5$ and correlation coefficient c, see Mr. Mittmann’s talk for post-processing of such bits

Why these RNGs? What do they have in common? (2)

- RNG 3: Quadrature measurement

$$|0\rangle = \int_{-\infty}^{\infty} \psi(x)\,|x\rangle\,dx$$

- "The quadrature measurement is conducted with a homodyne detector as shown in Fig. 1a. In such a detection system a weak signal (here the vacuum state) and a strong laser beam, called the local oscillator (LO), interfere on a symmetric beamsplitter to form two output beams with balanced powers. The two outputs are measured with two intensity detectors with carefully balanced amplifications, and the resulting electrical currents are digitized, subtracted and fed into a storage element. The difference current is proportional to the quadrature amplitudes of the vacuum state."

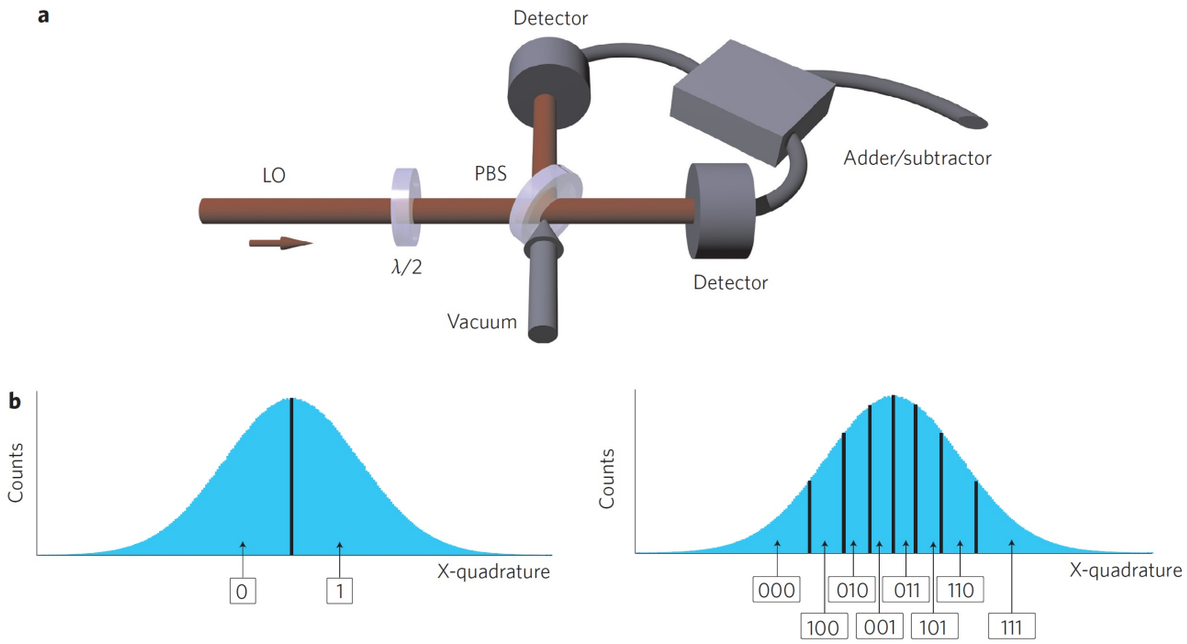


Figure 1

⟹ RNG 2 and RNG 3 have an approximation of a normal distribution as probability distribution function, RNG 1 would have it, if discretized in amplitude direction.

I. Digitization — from normal distribution to uniform distribution

a) Binning

- "Unbiased numbers . . . can be obtained by binning the measurement outcomes such that the integrated probability associated with each bin is equalized; that is,

$$\int_{-\infty}^{x_1} |\psi(x)|^2\, dx = \int_{x_1}^{x_2} |\psi(x)|^2 dx = \cdots = \int_{x_l}^{\infty} |\psi(x)|^2\, dx$$

  where $l+1$ is the number of bins. All the measurement outcomes within one bin are assigned a fixed bit combination (Fig. 1b). The length of this bit combination depends on the number of bins; that is for $l+1=2^n$ bins, the length of the bit combination is n."
- In other words, equidistant spacing of the cumulative distribution function.
- In their experiments, Gabriel et al. used 499968 sample points and $n=5$, i. e., 32 bins. They mention even the "advanced multilevel strategy process", i. e., Peres (von Neumann iteratively applied).
- Remark: If the empirical distribution function is not a perfect normal distribution[1] or it shows peaks, e. g., due to the non-linearity of the A/D converter, then one has errors (bias in the bits) in this discretization process.

[1] In reality, it has a Binomial distribution and only in the limit case a normal distribution.

Digitization — from normal distribution to uniform distribution (2)

b) Generalized von Neumann

- Generalized von Neumann procedure can be considered as part of digitization, cf. Bergmann generator with A/D converter in amplitude direction or R&S SIT TRNG.

- Primary effect: If the noise signal or the sample values have a normal distribution or an Binomial distribution, are independent and biased, then the bits after Generalized von Neumann are independent and unbiased.

- Secondary effects:
  - If we have peaks in the normal / Binomial distribution, then they are quadratically damped in the differences of sample values / after Generalized von Neumann.
  - The differences of sample values / the Generalized von Neumann procedure acts as a high-pass filter. For a sampling frequency of, e. g., $f_a = 50\,\mathrm{kHz}$, we have a low-cut frequency of approx. $2\,\mathrm{kHz}$, i. e., low frequency perturbations are filtered out.
  - N.B.: Differencing in time series analysis has often the effect to make a time series stationary.

von Neumann procedure (from 2018 talk)

- One of the oldest post-processing techniques
- J. von Neumann: *Various techniques used in connection with random digits*. 1951.
- Let $X_1, X_2, \ldots$ be binary random variables with realizations $b_1, b_2, \ldots$. Assume that X_i are independent and identically distributed (i.i.d.), but biased, i. e. $P(X_i = 1) := p$, $P(X_i = 0) := q = 1 - p$ with $0 \leq p, q \leq 1$. The procedure

(1)
$$\tilde{b}_j := \begin{cases} 0 & \text{if bit sequence } 01, \\ 1 & \text{if bit sequence } 10, \\ - & \text{else.} \end{cases}$$

  generates from n independent biased bits b_i approximately npq independent unbiased bits $\tilde{b}_j$.
- von Neumann outputs bits at irregular intervals. This is inevitable.

> *An algorithm for post-processing biased, but statistically independent random bits with a bounded number of input bits for one output bit cannot produce unbiased output bits for an infinite set of biases.*
>
> *M. Dichtl*

Generalized von Neumann = Peres procedure (from 2018 talk)

- The expected output rate at best (for unbiased and independent bits) of von Neumann procedure is $1/4$. How to improve this?

- Yuval Peres: *Iterating von Neumann's procedure for extracting random bits*. 1992.
- Let $Z_1, Z_2, \ldots$ be random variables that model sample values $z_1, z_2, \ldots, z_i \in \mathbb{R}^k$ with $k \geq 1$. Assume that Z_i are i.i.d. The procedure

(2)
$$b_j := \begin{cases} 0 & \text{if } z_{2i} < z_{2i+1}, \\ 1 & \text{if } z_{2i} > z_{2i+1}, \\ - & \text{else.} \end{cases}$$

generates independent unbiased output bits b_j. From n uniformly distributed sample values z_i we get approximately $\frac{n}{2} \times \frac{2^k - 1}{2^k}$ bits b_j. For $k = 1$ we obtain von Neumann's procedure.

Lemma (W. Killmann, Telekom Security)

Let X and Y be independent and identically distributed discrete random variables, which assume n different values. Then we have $P(\{X < Y\}) = P(\{X > Y\})$.

Proof: Let $V, V := \{v_i \mid i \in \overline{0, n-1}\}$, be the range of values that both random variables X and Y can assume. Then

$$\begin{aligned} P(\{X < Y\}) &= \sum_{i=0}^{n-1} P(\{X < v_i \mid Y = v_i\}) \cdot P(\{Y = v_i\}), \\ &= \sum_{i=0}^{n-1} P(\{X < v_i\}) \cdot P(\{Y = v_i\}) \quad \text{(independence of } X \text{ and } Y\text{)} \\ &= \sum_{i=0}^{n-1} P(\{Y < v_i\}) \cdot P(\{X = v_i\}) \quad \text{(identical distributions).} \end{aligned}$$

By applying the transformations backwards, we have

$$P(\{X < Y\}) = \sum_{i=0}^{n-1} P(\{Y < v_i\}) \cdot P(\{X = v_i\}) = \sum_{i=0}^{n-1} P(\{Y < v_i \mid X = v_i\}) \cdot P(\{X = v_i\}) = P(\{Y < X\}).$$

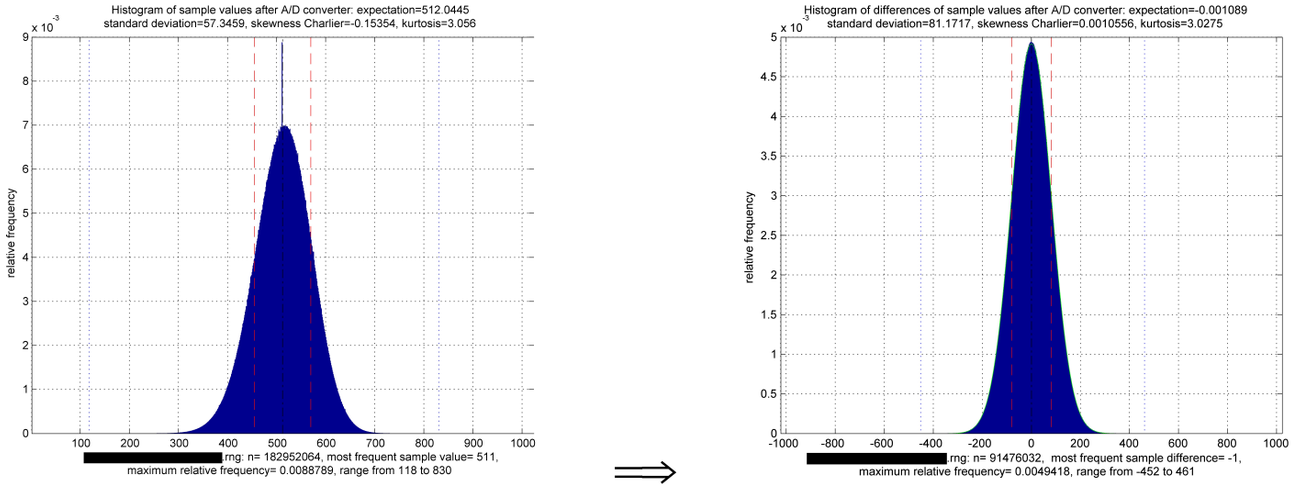
Remarks

(i) If $n = 2$, $X, Y \in \{0, 1\}$, then we have von Neumann procedure (1) using (2) for binary sources. Lemma can be generalized for discrete random variables which assume infinite many values and for continuous random variables.

(ii) Applying (2) to white noise, we have uniformly distributed independent bits: If the process $Z(t)$ is strongly stationary and the random variables are independent, then it follows from Lemma that the random variables $(B_j)_{j=0,1,2,\ldots}$ are uniformly distributed and independent.

(iii) GvN is independent from concrete distribution of discrete random variable.

(iv) Integral Non-Linearity of k-bit SAR (Successive Approximation Register) A/D converter leads to peak in histogram distribution of sample values at 2^{k-1} (and $2^{k-1} \pm 2^{k-2}$, ...): quadratic damping of peak in differences of sample values.



Analysis of peak: influence of differencing

Let $f(x;\mu,\sigma^2) := \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$ be the probability distribution function of $\mathcal{N}(\mu,\sigma^2)$ and δ be the Dirac functional. Approximate Dirac functional by limit sequence of normal distributions

$$\delta(x) = \lim_{\epsilon\to+0}\frac{1}{\sqrt{\pi}\epsilon}\exp\left(-\left(\frac{x}{\epsilon}\right)^2\right) = \lim_{\sigma_\epsilon\to 0}\frac{1}{\sqrt{2\pi}\sigma_\epsilon}\exp\left(-\frac{x^2}{2\sigma_\epsilon^2}\right) \quad \text{with } \epsilon^2 = 2\sigma_\epsilon^2.$$

Model of peak by ADC: $\alpha\mathcal{N}(\mu = 512, \sigma_\epsilon^2)$ $(k = 10!)$

Model of Z_{2i}: $X = Z_{2i} = \alpha\mathcal{N}(\mu = 512, \sigma_\epsilon^2) + (1-\alpha)\mathcal{N}(\mu,\sigma^2)$

Model of $-Z_{2i-1}$: $Y = -Z_{2i-1} = \alpha\mathcal{N}(-\mu = 512, \sigma_\epsilon^2) + (1-\alpha)\mathcal{N}(-\mu,\sigma^2)$

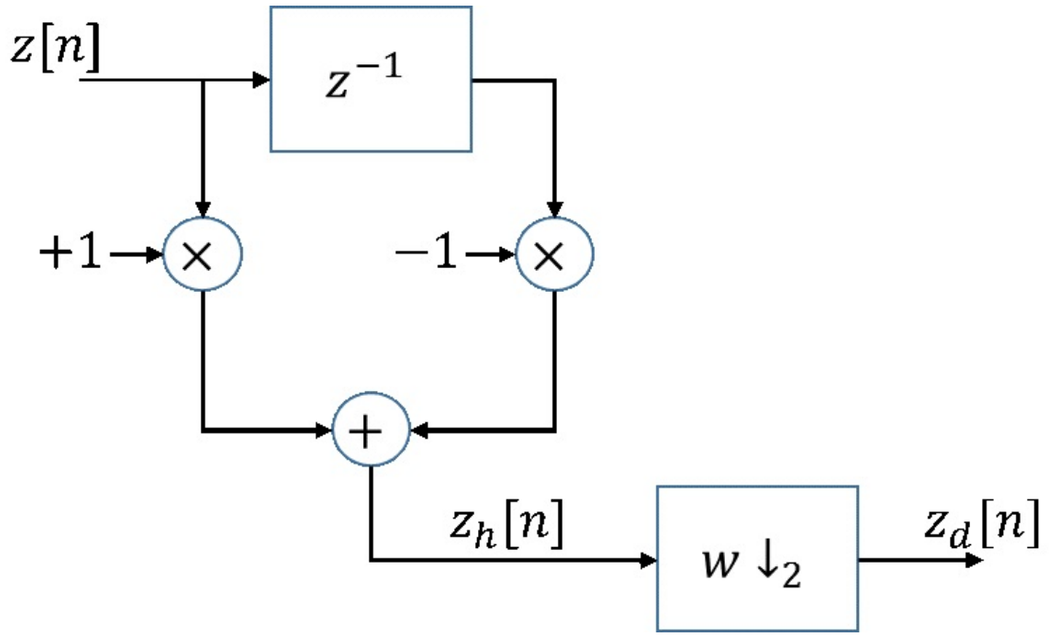
For independent random variables X and Y we have p.d.f. of $Z = X + Y$ by convolution $f_X * f_Y$, i. e., $f_Z(z) = \int f_Y(z-x)f_X(x)\,dx$. So, $Z \sim \mathcal{N}(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$, if $X \sim \mathcal{N}(\mu_X,\sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y,\sigma_Y^2)$ and X, Y independent. Assume $X = Z_{2i}$ and $Y = -Z_{2i-1}$ are independent, we have for $Z = X + Y$,

$$Z = \underbrace{\alpha^2\mathcal{N}(0,2\sigma_\epsilon^2)}_{\text{quadratic damping}} + \alpha(1-\alpha)\mathcal{N}(512-\mu,\sigma^2+\sigma_\epsilon^2) + \alpha(1-\alpha)\mathcal{N}(\mu-512,\sigma^2+\sigma_\epsilon^2) + \underbrace{(1-\alpha^2)\mathcal{N}(0,2\sigma^2)}_{\text{wanted signal}}$$

For $\sigma_\epsilon \to 0$ and $0 < \alpha < 1$ we have the quadratic damping effect (static or stationary view).

Implicit filtering (differences of sample values), Credits: F. Monsees, OHB

- Signal theoretic thoughts about filtering
  $z_d[n] = z[2n] - z[2n-1]$; $z[n]$ original sampled sequence, $f_a = 50\,\mathrm{kHz}$
- z^{-1} input delayed by one sample; high-pass filtering by $[+1, -1]$; downsampling by factor 2:
  $z_d[n] = z_h[2n]$



- Power spectral density by Wiener-Lee

$$S_{z_h,z_h}\left(e^{j\Omega}\right) = S_{z,z}\left(e^{j\Omega}\right) \cdot \left|H\left(e^{j\Omega}\right)\right|^2$$

with $\left|H\left(e^{j\Omega}\right)\right|^2$ frequency response of high-pass; $S_{z,z}\left(e^{j\Omega}\right)$ power spectral density of sequence $z[n]$

- We assume $S_{z,z}\left(e^{j\Omega}\right) = 1$, i. e., white noise.
- Frequency response of high-pass by Z-transform of impulse response $H(Z) = 1 - z^{-1}$. With $z = e^{j\Omega}$ we have

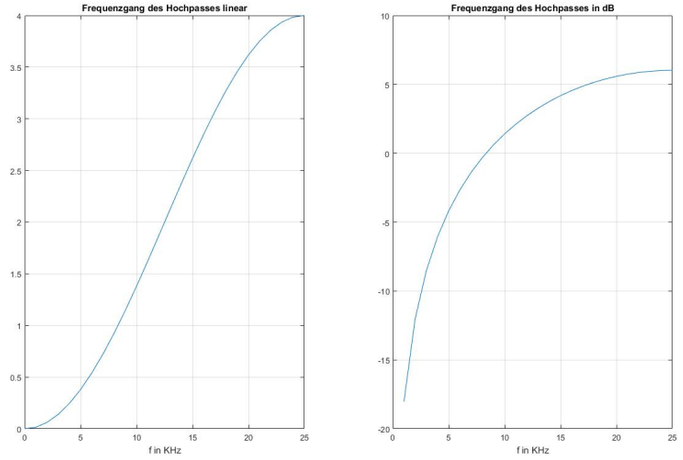
$$H\left(e^{j\Omega}\right) = 1 - e^{-j\Omega}.$$

$$\begin{aligned}\left|H\left(e^{j\Omega}\right)\right|^2 &= \left(1 - e^{-j\Omega}\right)\left(1 - e^{j\Omega}\right) \\ &= 2 - 2\cos(\Omega).\end{aligned}$$

So we have for the spectrum of signal $z_h[n]$

$$S_{z_h,z_h}\left(e^{j\Omega}\right) = S_{z,z}\left(e^{j\Omega}\right)\left[2 - 2\cos(\Omega)\right].$$

Implicit filtering (differences of sample values) (2)



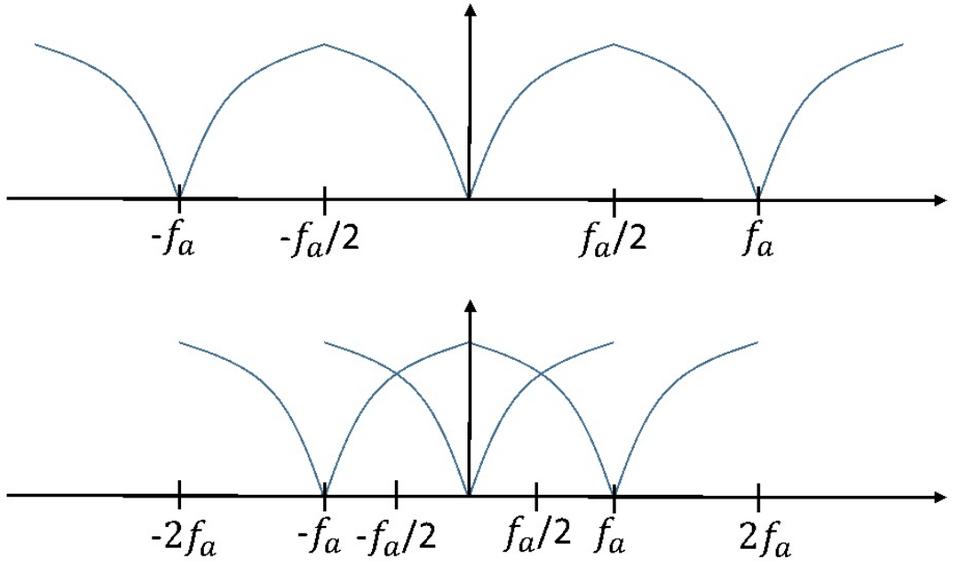
Frequency of high-pass, from 0 to $f_a/2$

- $\approx 2\,\mathrm{kHz}$ cut-off frequency
- Consider Wiener-Lee relation for ACF

$$r_{z_h,z_h}[\tau] = r_{z,z}[\tau] * r^E_{h,h}[\tau].$$

- With $h[n] = [1, -1]$ we have discrete energy-ACF $r^E_{h,h}[\tau] = h[n] * h[n] = [-1, 2, -1]$. $\Longrightarrow$ ACF of $r_{z_h,z_h}[\tau]$ will be widened.

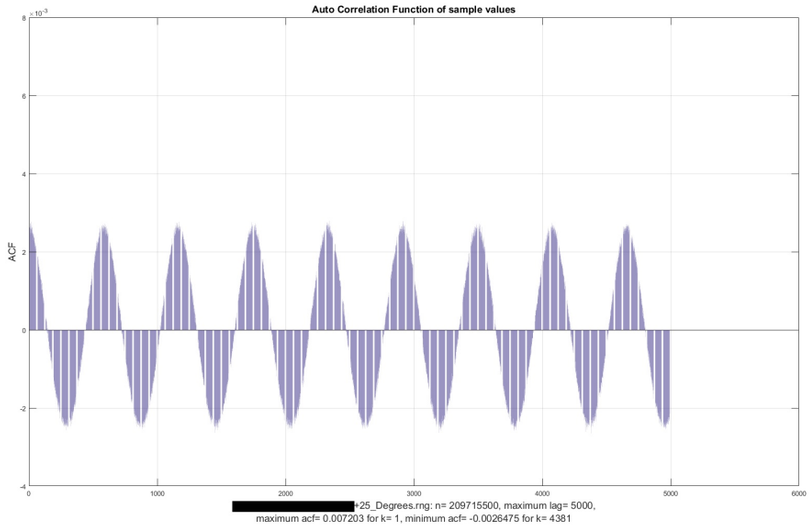
- Downsampling of sequence $z_h[n]$: periodic extensions of $S_{z_h,z_h}(e^{j\Omega})$ repeat with $f_a/2$ instead of f_a



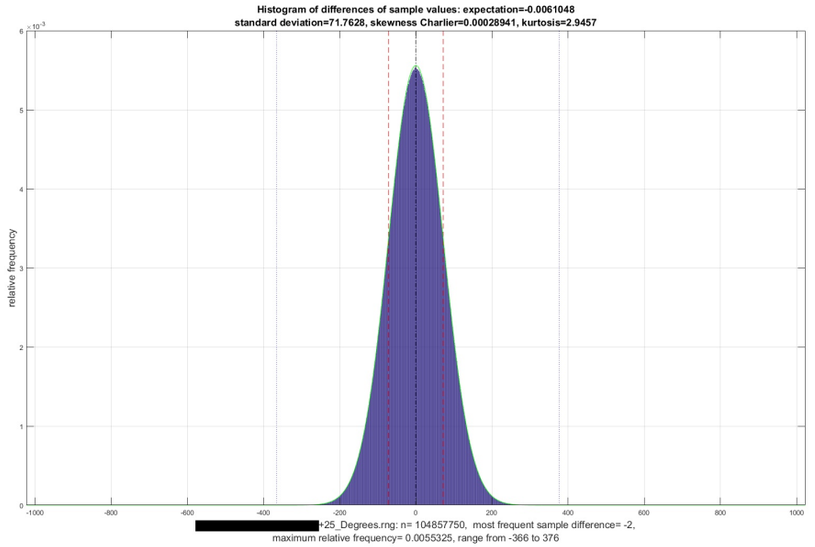
Influence of downsampling on psd of $S_{z_h,z_h}(e^{j\Omega})$: upper — before, lower — after downsampling

- narrow band, low frequency perturbations will be filtered by high-pass

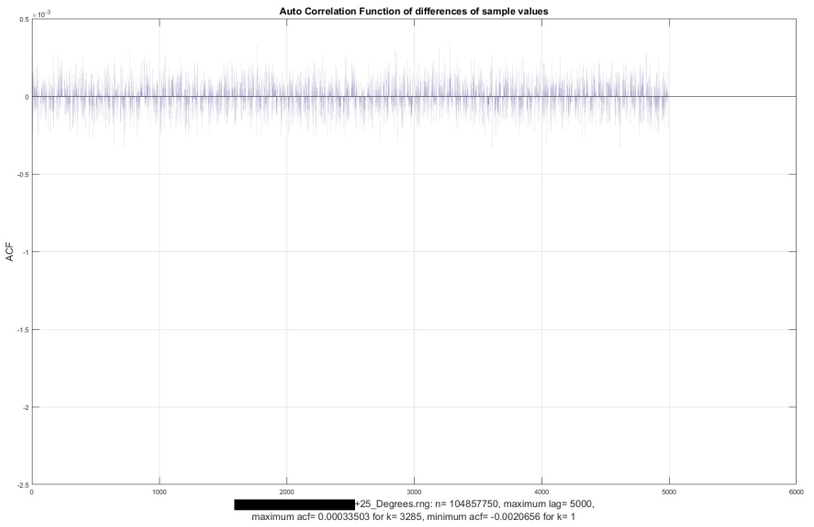
Implicit filtering (differences of sample values) (3) — Example



ACF of sample values periodic, then sample values periodic. Houston we have a problem!



Histogram of differences of sample values looks perfect



ACF of differences of sample values looks good

- Explanation: Missing terminating resistor in measurement setup for radiation tests (expensive, cannot be repeated easily) leads to $50\,\mathrm{Hz}$ signal
- Differences of sample values / Generalized von Neumann acts as high-pass filter and filters out $50\,\mathrm{Hz}$ perturbation

II. Algorithmic or mathematical post-processing

- algorithmic post-processing, ~~forthcoming~~ BSI AIS 20/31: "A type of post-processing that is generally used for the purpose of increasing the entropy per data bit (entropy extraction). It is usually applied to the raw random numbers. The name is chosen to distinguish it from an analog transformation (e. g., amplification, band-pass filter).

  Note 1: Viewed as a mathematical function, algorithmic post-processing algorithms usually have small domains and small ranges (in contrast to cryptographic post-processing). Algorithmic post-processing can be stateful (i. e., with memory) or stateless.
  Note 2: Typical examples of algorithmic post-processing algorithms: XORing bits or binary vectors, modular addition, LFSRs."
- Examples of non-cryptographic post-processing:
  - XOR, von Neumann's method, length of runs method, Generalized von Neumann/Peres procedure, Optimal XOR-constructions, $[n, m, t]$-resilient functions, ...
  - (Randomness extractors in QRNGs: Toeplitz-hashing extractors, Trevisan's extractor, ...)?

We consider now c) XOR and d) von Neumann procedure

XOR and Piling-Up Lemma (from 2020 talk)

- Let $X_1, \ldots, X_n$ be i.i.d. binary random variables with $P(X_i = 1) := p$ and $\epsilon := p - 0.5$ (bias). Let $Y := X_1 \oplus \cdots \oplus X_n$. What is the bias of Y?
- For $n = 2$ we obtain

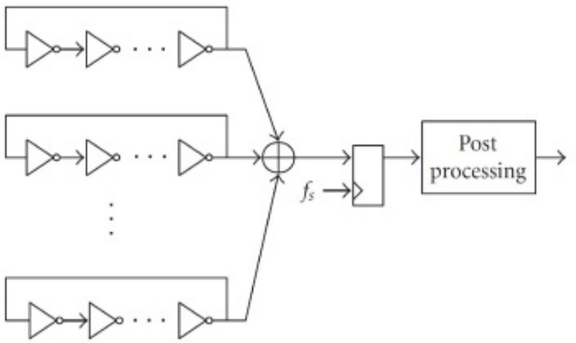
$$\varepsilon_{\text{total}} = P(X_1 \oplus X_2 = 1) - P(X_1 \oplus X_2 = 0) = -4\epsilon^2$$

- For arbitrary n we obtain the Piling-up Lemma

$$\varepsilon_{\text{total}} = P(X_1 \oplus \cdots \oplus X_n = 1) - P(X_1 \oplus \cdots \oplus X_n = 0) = (-1)^{n+1}(2\epsilon)^n$$

- $\Longrightarrow$ The bias decreases exponentially by XOR-ing n independent variables. Thus, XOR-ing random bits is used for unbiasing.
- This may be useful for $n = 2$ or $n = 3$. But for large n, for example $n = 114$ it becomes dangerous.

Too many XORs



- B. Sunar, W. Martin, D. Stinson: *A provably secure true random number generator with built-in tolerance to active attacks*, IEEE Trans. Computers 56(1), 2007.
- 114 ring oscillators, assumed independent and identically distributed, XORed in huge binary XOR tree
- post-processing is fine, resilient function

- Markus Dichtl et al. showed in a number of papers [2007, 2008] that this RNG is broken:
  1. unrealistic assumptions on timing for XOR tree
  2. independence assumption is regularly not fulfilled (this is used for RO PUFs)
- K. Wold, C.H. Tan [2009] fixed the first problem by introducing a D-flip-flop after every ring oscillator[1]. They noted, that now the resilient function is no longer necessary!
- Independence can be somewhat mitigated by placing ring oscillators at different areas of the chip

[1] This should also fix early propagation problems, see side channel-analysis.

What to do with dependent bits?

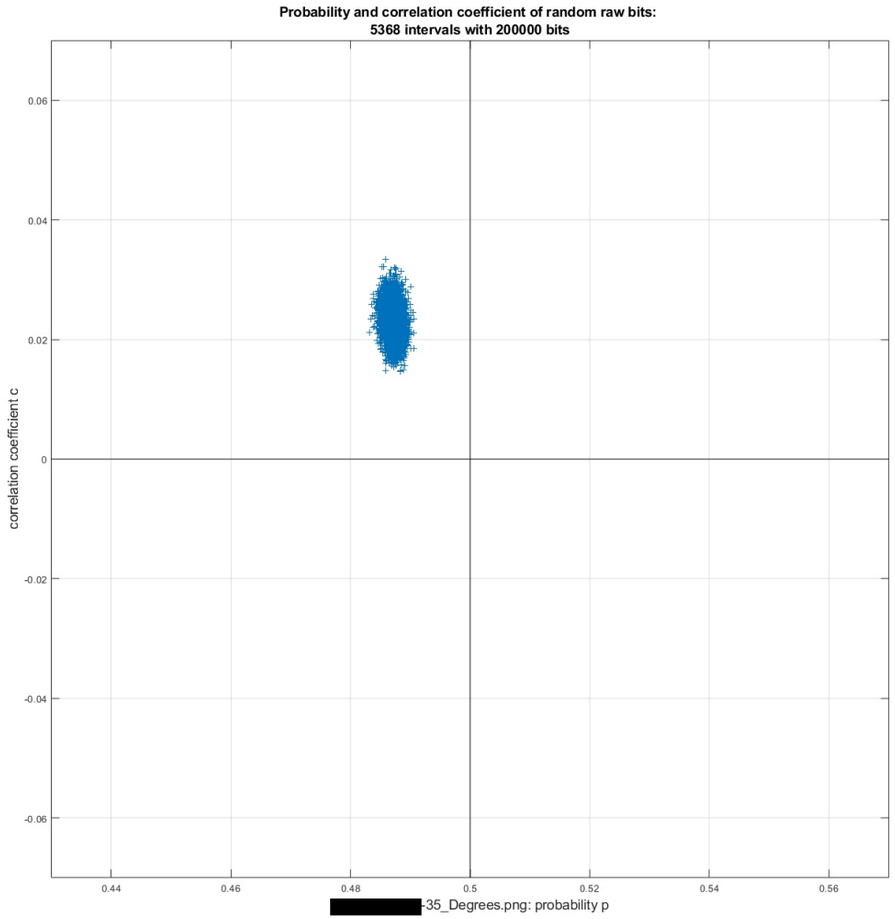
- RNG evaluation: Hardware problem (power supply) lead to dependencies under extreme conditions ($-35\,$°C). The entropy source itself is fine (with other power supply).
- Can we do something about the large dependency?
- Stochastic model with probability p (or bias $p - 0.5$) and auto-correlation / serial correlation c
- Parameter estimation for binary random variables $X_1, X_2, \ldots$:

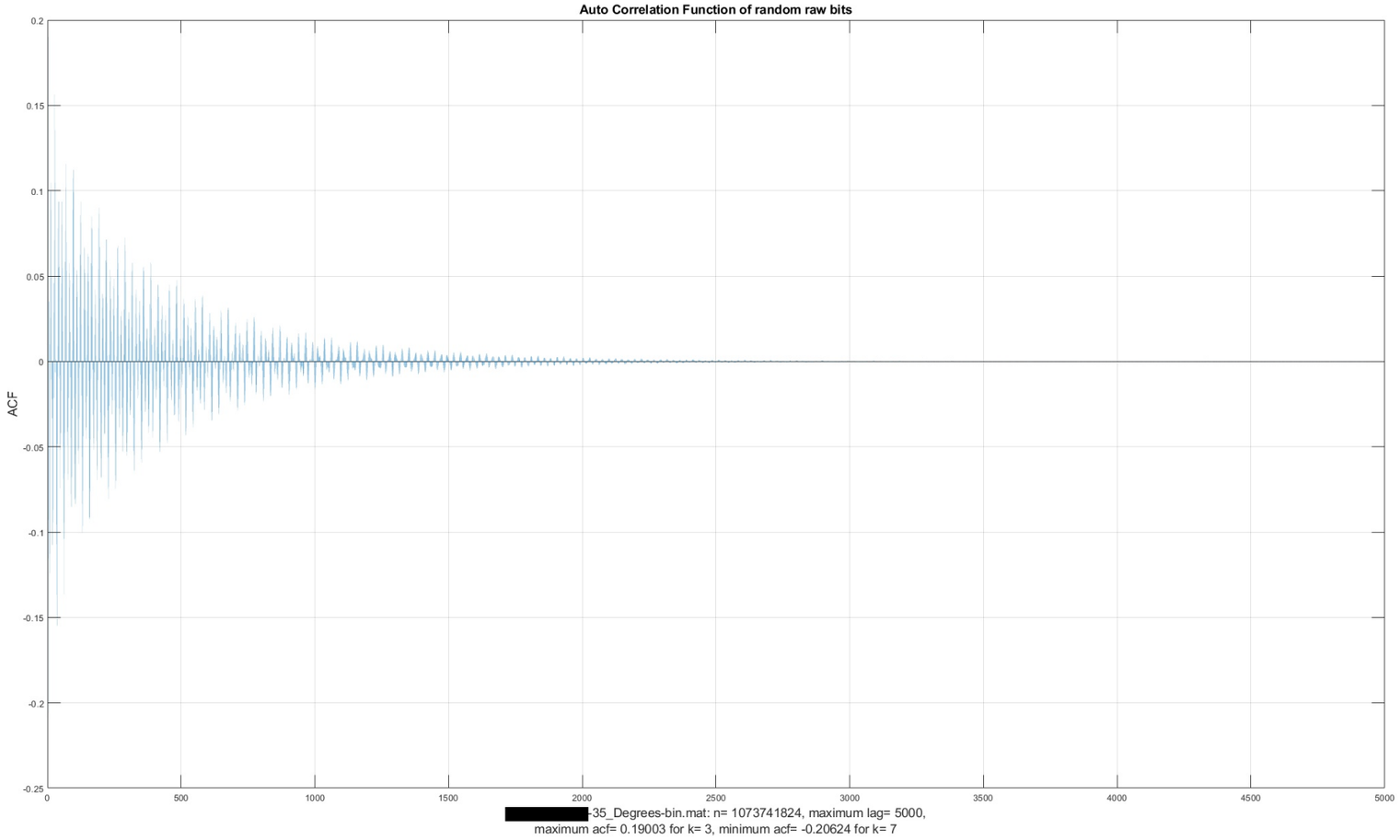
(3a)
$$\hat{p} = \frac{\sum_{j=1}^{n} X_j}{n} = \frac{n_1}{n} = \bar{X},$$

(3b)
$$\hat{c} = \frac{\frac{1}{n}\sum_{j=1}^{n-1} X_j X_{j+1} - \bar{X}^2}{\frac{1}{n}\sum_{j=1}^{n} (X_j - \bar{X})^2}.$$

- We have 2^{30} bits. We estimate parameters on 5368 intervals with 200000 bits each. The results are, well, not so good. Old AIS 20/31 bounds are $|\hat{p} - 0.5| \leq 0.025$ and $|\hat{c}| \leq 0.02$ for approx. 200000 bits.

Parameter estimation (bias, correlation) and ACF: original data

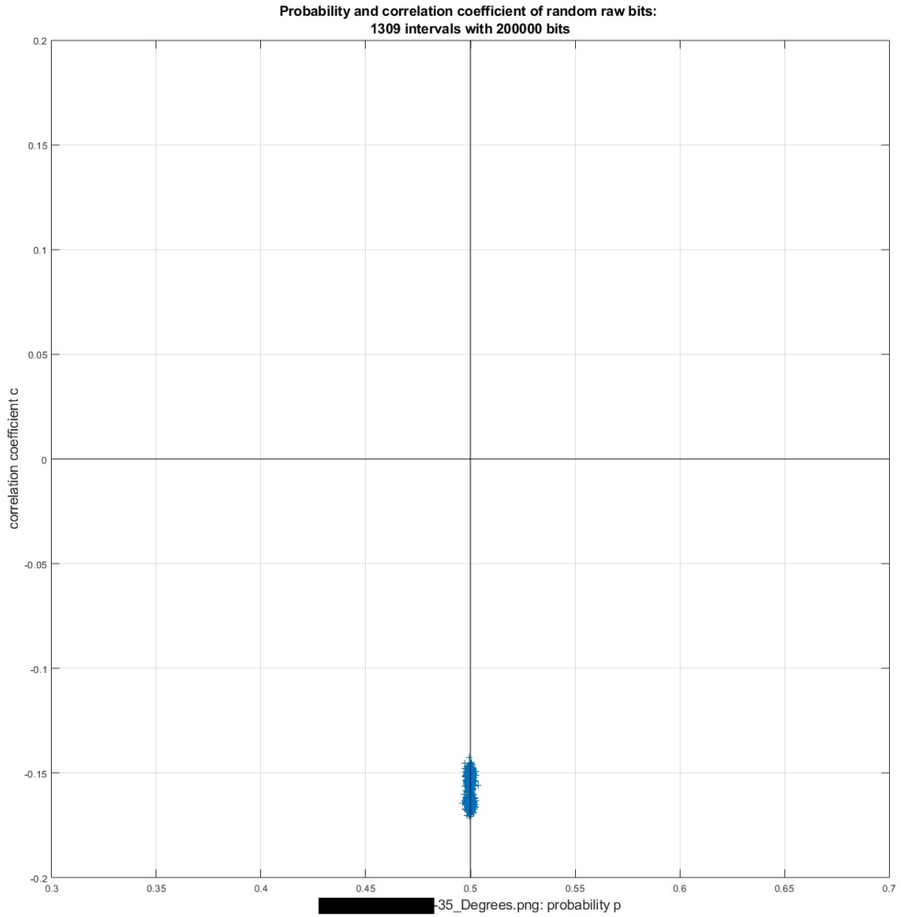


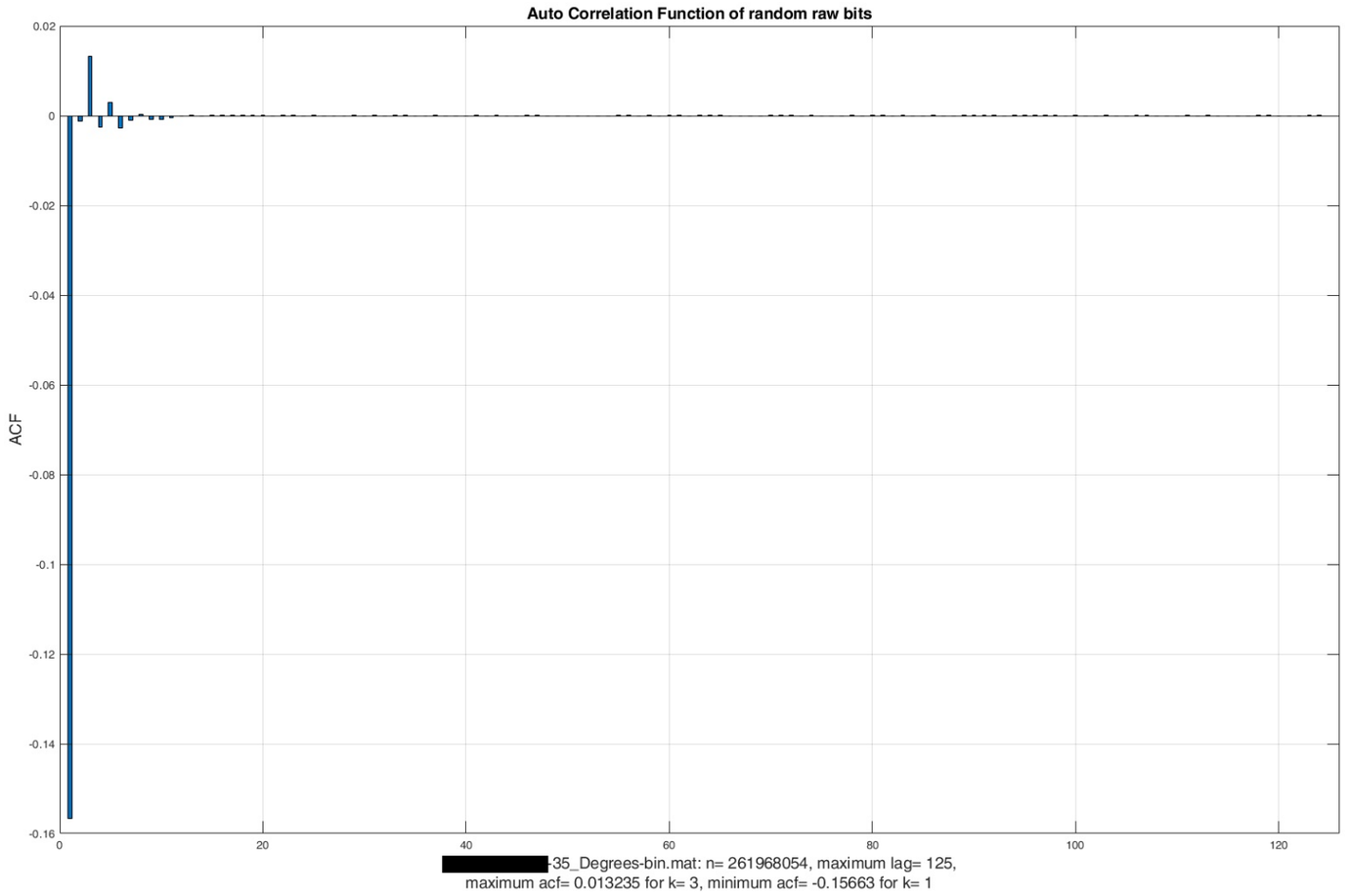


ok: $-0.0167 \leq p - 0.5 \leq -0.0094$,

bad: $0.0146 \leq c \leq 0.0334$

Parameter estimation (bias, correlation) and ACF: after von Neumann

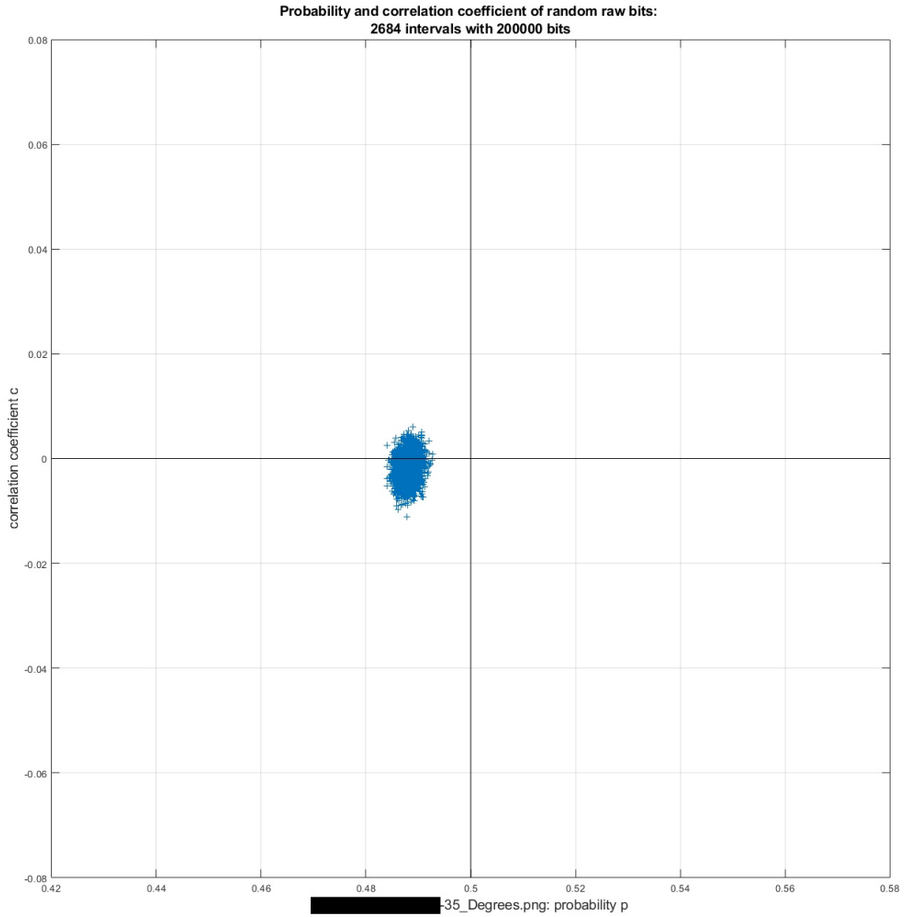


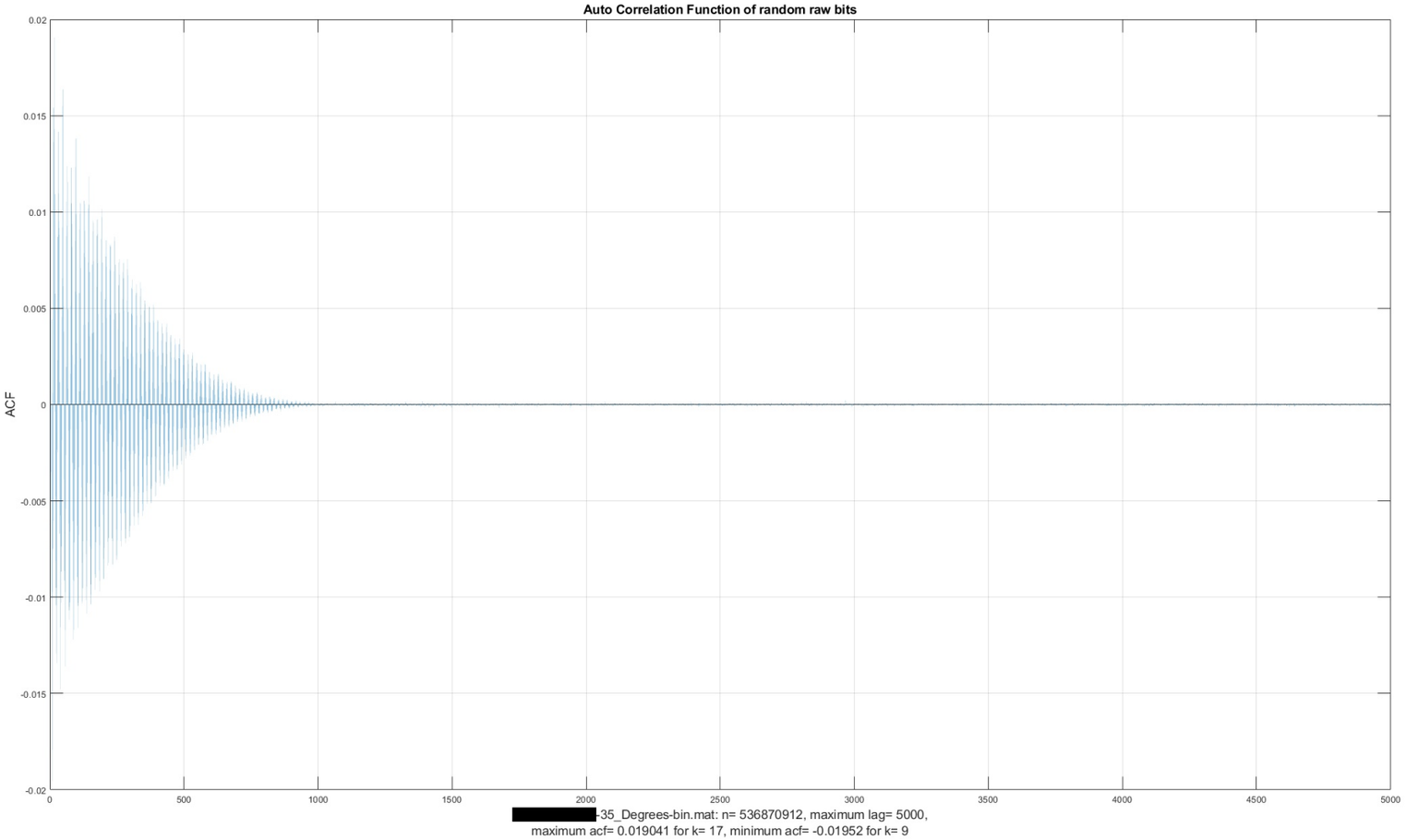


good: $-0.0037 \leq p - 0.5 \leq 0.0035,$

bad: $-0.170 \leq c \leq -0.142$

Parameter estimation (bias, correlation) and ACF: after XOR of two bits

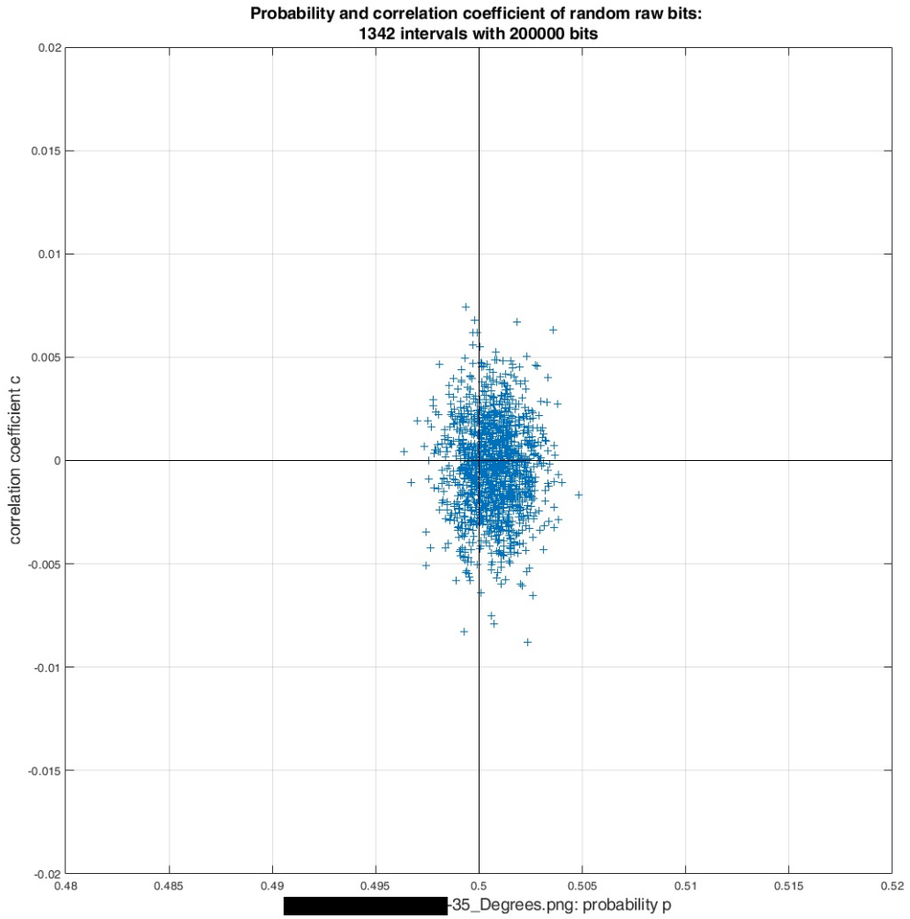


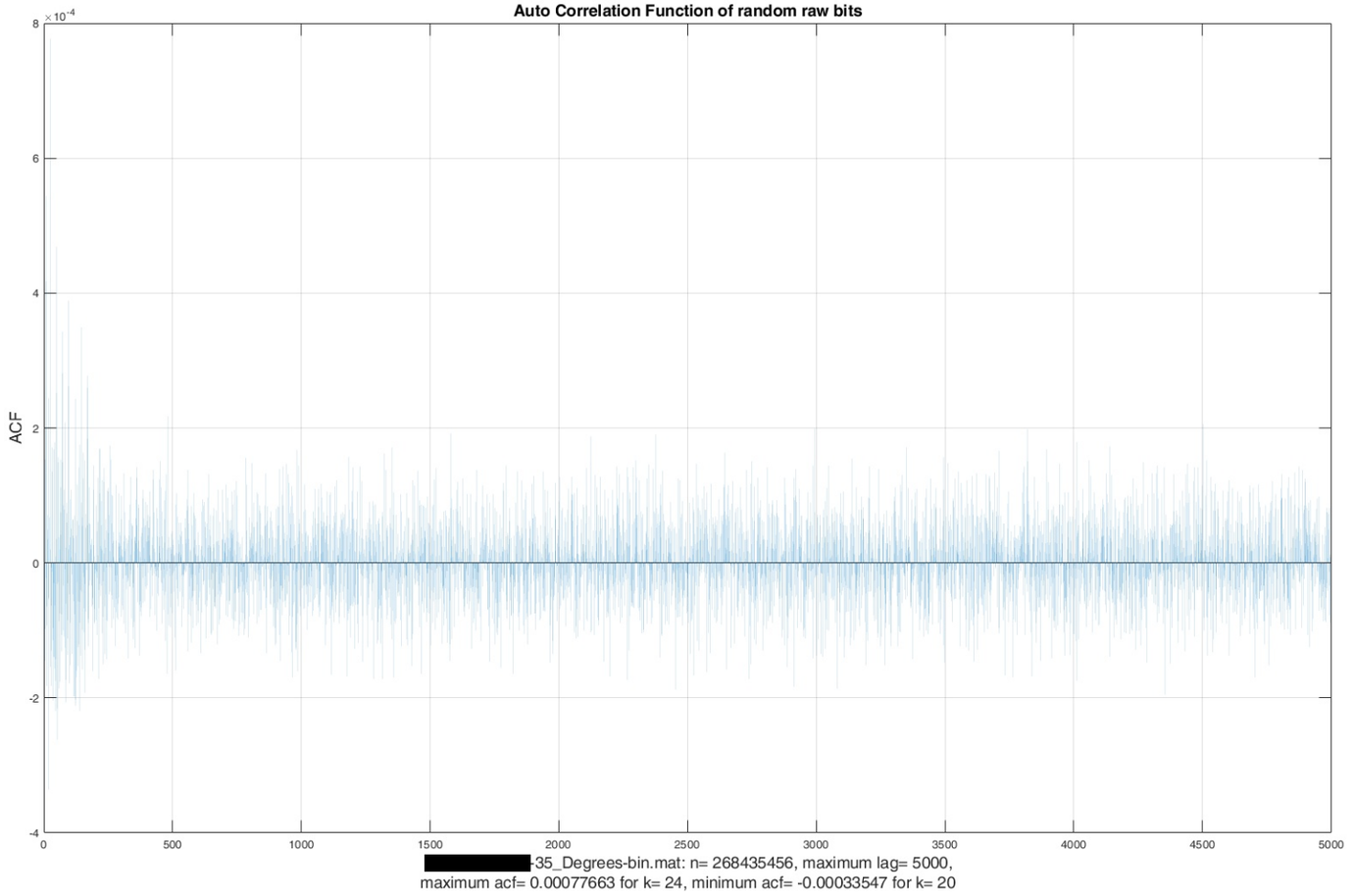


ok: $-0.0159 \leq p - 0.5 \leq -0.0072$,

good: $-0.011 \leq c \leq 0.006$

Parameter estimation (bias, correlation) and ACF: after XOR of four bits





good: $-0.0036 \leq p - 0.5 \leq 0.0048,$

good: $-0.0087 \leq c \leq 0.0074$

Summary

- For independent bits, von Neumann is superior to XOR of two or four bits as it removes the bias completely.
- For the dependent and biased bits in our example, von Neumann gives bad results as it cannot remove / reduce the dependency / correlation.
- XOR of four bits and von Neumann give approximately the same number of bits.
- Only XOR of four bits gives good results for bias and dependency!

Our experimental results for perturbed or dependent random bits

I. Digitization	II. Mathematical post-processing
a) Binning: requires rather exact cumulative distribution function	c) XOR: XOR of four bits removes dependency and bias
b) Generalized von Neumann: very stable; quadratic damping of peaks; high-pass filtering of low frequency perturbations	d) von Neumann: no satisfactory results for dependent and biased bits

Summary (2)

- If we assume X_i as Markovian with two parameters, namely (bias, correlation) = one-step dependent Bernoulli experiments, the XOR of X_i and X_{i+1} is not longer Markovian, see W. Schindler: *Evaluation Criteria for Physical Random Number Generators*, Example 3.9. In: *Cryptographic Engineering*. Ed. by Ç. K. Koç, Springer 2009. pp. 25–54.
- The exact distribution is rather difficult, see H. J. Helgert: *On sums of random variables defined in a two-state Markov chain*, J. Appl. Prob. **7**, 761-765 (1970).
- see talk by J. Mittmann: *Post-processing algorithms for Markov chain models*
- What's missing after these talks? Online tests for biased and dependent bits

. . . the rest
Shall bear the business in some other fight

Shakespeare, Coriolanus

The rest is silence

Shakespeare, Hamlet