Maximization of random bits extraction rate using a digitalized parallel quantum random number generator

Mauricio Ferreira[1,2], Nuno A. Silva[1], Armando N. Pinto[1,2], and Nelson J. Muga[1]

[1] *Instituto de Telecomunicações, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal*

[2] *Department of Electronics, Telecommunications and Informatics, University of Aveiro, 3810-193 Aveiro, Portuga*

A schematic representation of the experimentally implemented Quantum Random Number Generator (QRNG) is shown in Fig. 1. In practice, a 1550.92 nm continuous-wave laser tuned at approximately 11 dBm is used as the Local Oscillator (LO), while a Variable Optical Attenuator (VOA) (VOA1) allows accurate control of its output power. Moreover, an 80/20 Beam Splitter (BS) (BS1) and an Optical Power Meter (OPM) are introduced to monitor the input power at the 50/50 BS (BS2). This optical signal was measured at 5.5 mW. A second VOA (VOA2) was additionally used to fine-tune the balancing condition of the detection scheme. A balanced receiver (WL-BPD1GA) with an output bandwidth from 300 kHz to 1 GHz and a transimpedance gain of 3500 V/W is then introduced to detect the output signals. Finally, the output is sampled at 983.04 MSa/s by an Analog-to-digital Converter (ADC) module (Texas Instruments ADS54J60EVM) with a ±0.95 V acquisition range and resolution of 16 bits.

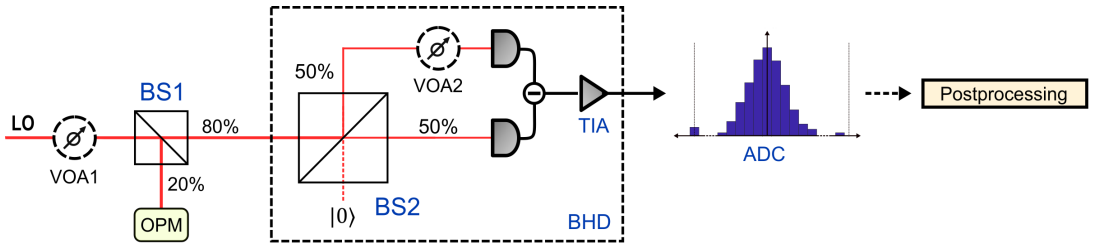


Fig. 1: Schematic of the experimental setup implemented for the proposed vacuum-based QRNG. The LO interacts with the vacuum state in a BS (BS2) with its second input port blocked. Its output signals are subsequently detected in a Balanced Homodyne Detection (BHD) scheme, and the resulting signal is amplified by a high-gain Transimpedance Amplifier (TIA) and digitized by a high-resolution ADC. Adapted from [1].

As specified by the Nyquist sampling theorem, the output signal is only correctly recovered when its bandwidth is less than half of the sampling frequency, f_s. Here, however, this rule should not be followed since, given a finite signal bandwidth, any samples taken would necessarily be highly correlated. In fact, under an ideal TIA response, maximally uncorrelated measurements are only obtained when the sampling frequency, f_s, follows [2]:

$$f_s = \frac{2\Delta f}{j}, \quad \forall j \in \mathbb{N}, \tag{1}$$

where Δf is the cut-off frequency of the detector's TIA. Consequently, although an optimal sampling rate was not used in this experimental implementation, a rate lower than 2 GSa/s should always be adopted to avoid introducing additional correlations in the output signal [3].

Assuming that all classical noise can be known to an adversary, approximately 8.39 bits is expected to be extracted from each ADC measurement [1]. At the chosen sampling rate, this allows generation rates up to 8.23 Gbps with a security parameter, ϵ, of 2^{-105}, despite the effective throughput being limited by the randomness extractor implemented and thus highly dependent on the hardware implementation. Nonetheless, a maximum theoretical entropy of 13.97 bits per sample could be extracted by amplifying the homodyne signal to an optimized level. Unfortunately, in practice, amplifying the homodyne signal would

deteriorate the observed Quantum-to-classical Noise Ratio (QCNR) due to the additional noise introduced by the amplifier. More critically, operational amplifiers are defined by their Gain-Bandwidth Products (GBWPs), and their effective bandwidth is inversely proportional to the chosen gain [4]. Consequently, given the restriction imposed by Equation (1), this would necessarily result in slower sampling rates, which could offset the effect of entropy optimization. Nevertheless, one could circumvent this problem by relying on a spectrally multiplexed QRNG scheme that amplifies non-overlapping sideband frequency modes of the vacuum state within the detector's bandwidth. These modes are independent and thus constitute subentropy sources, which generate parallel random generation channels that can be sampled at lower rates [5]. Given the previous experimental implementation, the potential performance of such a multiplexing scheme was briefly characterized by numerically simulating the ideal operation of the required electronic components in Matlab.

1 Multiplexed QRNG

A schematic representation of the proposed scheme can be seen in Fig 2. In practice, this stage is constituted by a Power Splitter (PS) that divides the output signal into an arbitrary number of channels. Each subsequent output is mixed down in a frequency mixer (M_N) with a signal whose frequency f_{ch} defines the center of the selected spectral sideband. Posteriorly, the channel is filtered through a Low-pass Filter (LPF) with a certain cutoff frequency f_c, defining its bandwidth. Finally, each channel is amplified to the corresponding optimization level by two amplifiers (A_{N1} and A_{N2}) of equal gain, G_{amp} , that are, respectively, applied before and after the spectral selection. This two-stage amplification approach requires lower GBWPs to be considered for each amplifier.

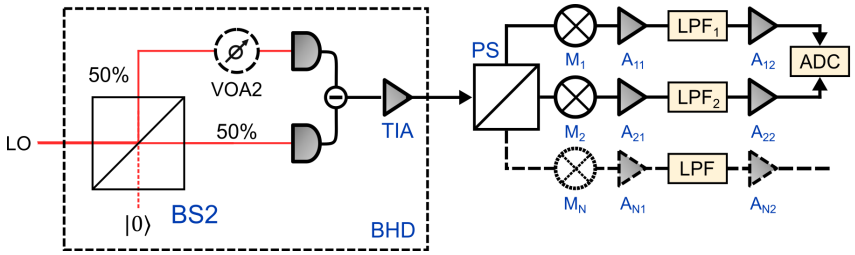


Fig. 2: The homodyne output signal is divided by a PS into an arbitrary number of non-overlapping channels. Each frequency sidemode is posteriorly selected by a frequency mixer (M_N), filtered by an LPF, and posteriorly amplified to optimize the available entropy in each channel. Adapted from [6].

Here, two parallel randomness generation channels were considered, given the dual-channel ADC that was selected for the experimental implementation. The first channel was mixed down with a 200 MHz sinusoidal wave, while the second component was used to sweep central frequencies up to 600 MHz. Moreover, an almost-ideal 40th-order Butterworth Infinite Impulse Response (IIR) LPF with f_c of 100 MHz was initially implemented, and a low-noise operational amplifier with a noise power density of $0.98\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ was selected to implement the two-stage amplifier. Here, the effective bandwidth of each amplifier was determined by a similar 4th-order LPF filter. For all other parameters, the operation of the electronic components was considered ideal.

Figs. 3a and 3b show the conditional min-entropy as a function of the gain of each individual amplifier stage, G_{amp}, for, respectively, GBWPs of 1 GHz and 8 GHz. Here, note that the global amplification of this two-stage amplifier is twice the G_{amp} displayed. As can be seen, a gain of approximately 28 dB is required to reach entropy optimization. Meanwhile, only 26 dB are necessary if a GBWP of 8 GHz is considered. This difference is due to the bandwidth reduction imposed by the amplifiers for larger channel gains and can be seen when the spectral selection stage is removed. In this case, the effective bandwidth of the output signal is no longer limited to the 100 MHz cut-off frequency of the LPF, and its observed variance increases. Consequently, lower gain factors are required to reach the same entropy values. For a 1 GHz GBWP, this difference disappears for gains larger than 20 dB since this is the level at which the effective bandwidth imposed by the amplifiers equals f_c, and the two cases become equivalent. For higher amplification levels, the sideband width is instead defined by the effective bandwidth of the amplifier. Meanwhile, the same effect is not observed for an 8 GHz GBWP since, here, these bandwidths

would only become equal at approximately 38 dB. As such, amplifiers with high GBWPs are necessary despite this requirement potentially being minimized by introducing additional amplification stages.

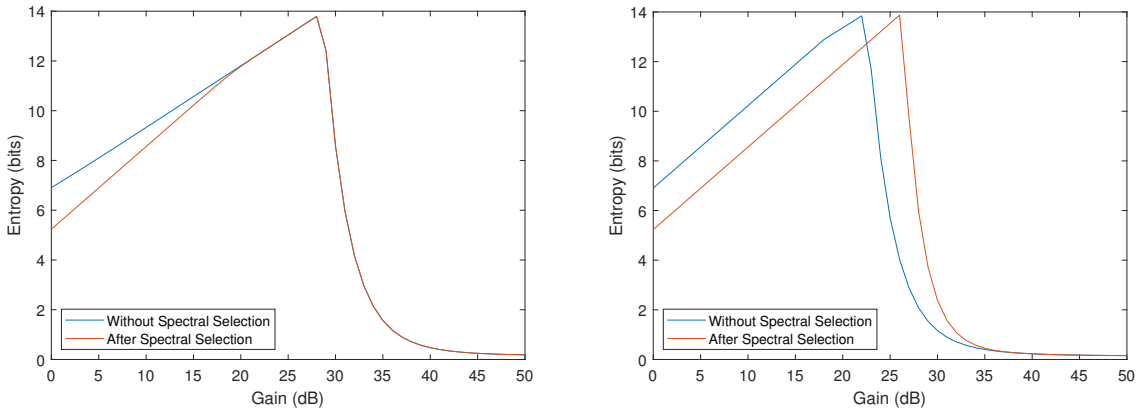


Fig. 3: Worst-case conditional min-entropy as a function of the gain of each amplifier for a GBWP of (a) 1 GHz and (b) 8 GHz.

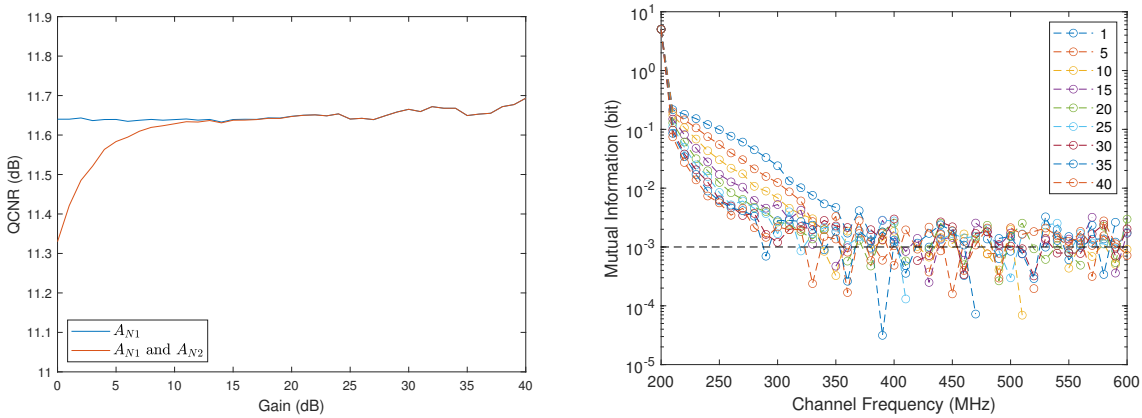
Nonetheless, to better assess the impact of these experimental limitations, the case of a 1 GHz GBWP will be considered in the following section. In fact, for the case of an 8 GHz product, the effective bandwidth does not limit the implementation for channel widths up to 400 MHz. As can be seen in Fig. 4a, a deterioration of the QCNR is expected, particularly due to the additional electronic noise introduced by the amplifiers. Here, a comparison with the case of an ideal noiseless second amplifier, A_{N2}, was made. It is observed that the impact of the excess noise from this particular amplifier is mainly limited to small amplification gain factors due to the low-noise amplifier considered. A maximum reduction of approximately 0.4 dB is expected in comparison with the QCNR of the non-multiplexed scheme. Nevertheless, as a consequence of the two-stage scheme explored, a small QCNR deterioration is still seen even when the excess noise of the second stage is not considered. For this, the noise from the first amplifier, A_{N1}, which is itself amplified by the second stage, plays a significant contribution.

With two sequences of 1 M samples, the mutual information was also calculated for increasing channel separation and LPF orders. The mutual information, which yields the amount of information gathered about one variable by measuring the other, between two continuous variables X and Y is defined as [7]:

$$I(X:Y) = \int_{y\in\mathcal{Y}} \int_{x\in\mathcal{X}} p_{X,Y}(x,y) \log_2 \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \; dx \; dy, \tag{2}$$

where $p_X(x)$, $p_Y(y)$ are the marginal probability density functions of each variable, and $p_{X,Y}(x,y)$ is their joint probability density function. Here, this quantity was obtained using the estimator based on k-nearest neighbor distances proposed by [8] and implemented by [9]. As can be seen in Fig. 4b, for close-to-ideal filters, this quantity rapidly converges to the estimator distribution expected for independent sequences of this size. Here, the black-dashed line represents the theoretical standard deviation expected. This seems to validate this approach as an adequate method to insert parallel Entropy Sources (ESs) within the homodyne detector bandwidth, with minimal changes needed in the experimental implementation. Despite this, as can be seen, for more realistic filters, side-information can still be present up to twice the channel bandwidth. This should be accounted for when designing a practical implementation, namely by carefully characterizing the electronic circuit. Ultimately, more than the chosen channel width, how much information is leaked by each ES will limit their allowable spectral proximity and, thus, the number of simultaneous channels that can be experimentally implemented.

Finally, the expected generation rate for each channel was assessed as a function of the chosen channel width. With this aim, f_c values of 200 MHz, 150 MHz, 100 MHz, and 50 MHz were considered. However, note that, for the analyzed 1 GHz GBWP and given the required 28 dB amplification gain, a maximum sampling rate of approximately 79.6 MSa/s is allowed at each channel. Consequently, although an average of 13.79 bits was obtained across all f_c values, the channel throughput is limited to 1.10 Gbps regardless of the chosen width. In Fig. 5a, this analysis is repeated for an amplifier with a GBWP of

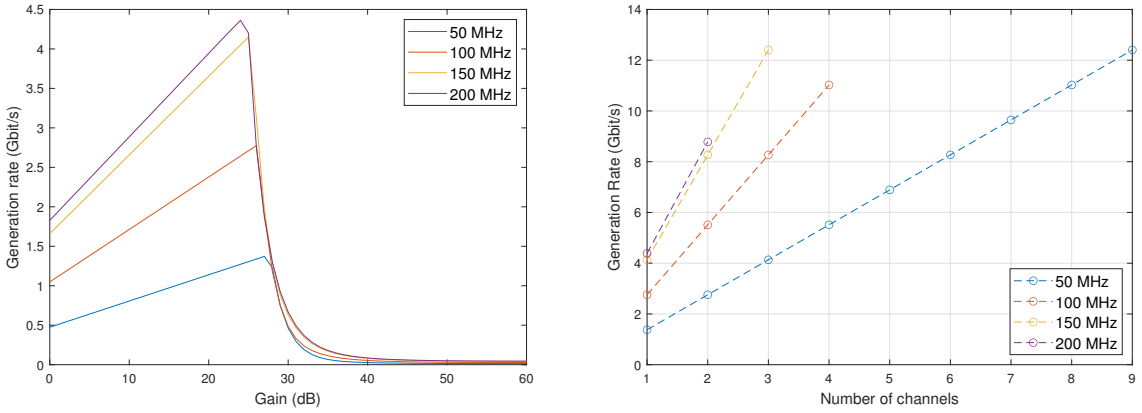


(a) (b)

Fig. 4: (a) QCNR as a function of the individual amplifier gain G_{amp}. This noise figure was evaluated both while considering only the excess electronic noise introduced by A_{N1} and by the two amplifiers with 1 GHz GBWP. (b) Mutual information between the two channels as a function of the f_{ch} from Channel 2. Channel 1 was fixed at f_{ch} =200 MHz, and sequences of 1 M samples were considered.

---

4 GHz and amplification levels up to 60 dB. As represented, when the GBWP is longer a limiting factor, smaller channel bandwidths also require slightly higher channel amplification levels to reach an optimized min-entropy. Nevertheless, again, an average of 13.79 bits bits was obtained across all possible channels. Correspondingly, the throughput increases proportionally with the maximum sampling rate. For this amplifier, this limit is now twice the f_c for all channels considered, except for the 200 MHz, which is limited by the effective amplifier bandwidth. This reflects itself as a correspondingly smaller generation rate increase for this particular case.



(a) (b)

Fig. 5: (a) Theoretically expected channel generation rate for different LPF cut-off frequencies as a function of the amplification level. (b) Expected maximum generation rate for a k-channel multiplexed QRNG at different LPF cut-off frequencies. In these two analyses, exclusively, a GBWP of 4 GHz was selected.

Following this analysis, as shown in Fig. 5b, a dual-channel configuration exploring the 4 GHz amplifier would allow generation rates up to 5.51 Gbps. Consequently, similarly to the case of the 1 GHz GBWP, no performance gain would actually be obtained. Regardless, this multiplexing technique could relax the performance requirements for the physical devices by lowering the required ADC sampling rates, which is important to reach low-cost high-speed implementations. Moreover, the chosen f_c of 100 MHz supports up to 4 channels within the detector's bandwidth. Thus, assuming that the extractable entropy remains constant between all channels, an additional third or fourth channel would increase this value to 8.27 Gbps and 11.02 Gbps, respectively. This corresponds, respectively, to increases of 0.49% and 33.9%

in comparison to the experimentally implemented non-multiplexed scheme described in the previous section. Alternatively, the channel bandwidth could be slightly increased to 150 MHz, which would allow a throughput of approximately 12.4 Gbps. This would constitute an increase of 50.7% in the supported maximum generation rate. Nonetheless, since the frequency sidebands must be properly spaced, some f_c values cannot support the maximum number of channels theoretically allowed. Without careful consideration of the values chosen, lower bandwidths appear to allow for better maximization of the overall generation rates since they lead to a smaller decrease in the overall throughput per channel not implemented. On the other hand, they require more channels to achieve the same output rates. It thus appears that an actual experimental characterization of the scheme would be necessary to determine their minimum secure distance and select the channel configuration. This will always be a compromise between higher generation rates and the feasibility of implementation.

References

[1] Maurício J. Ferreira, Nuno A. Silva, Armando N. Pinto, and Nelson J. Muga. Statistical validation of a physical prime random number generator based on quantum noise. *Applied Sciences*, 13(23), 2023.

[2] Yong Shen, Liang Tian, and Hongxin Zou. Practical quantum random number generator based on measuring the shot noise of vacuum states. *Phys. Rev. A*, 81:063814, Jun 2010.

[3] Ferreira, Maurício J. and Silva, Nuno A. and Pinto, Armando N. and Muga, Nelson J. Characterization of a quantum random number generator based on vacuum fluctuations. *Applied Sciences*, 11(16), 2021.

[4] Texas Instruments Incorporated. Noise analysis in operational amplifier circuits. Technical report, 2007. Accessed October 2022.

[5] Xiaomin Guo, Chen Cheng, Mingchuan Wu, et al. Parallel real-time quantum random number generator. *Opt. Lett.*, 44(22):5566–5569, Nov 2019.

[6] Nuno A. Silva, Maurício J. Ferreira, André Carvalho, et al. A network server for distributing quantum random numbers. In *2023 23rd International Conference on Transparent Optical Networks (ICTON)*, pages 1–4, 2023.

[7] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006.

[8] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Phys. Rev. E*, 69:066138, Jun 2004.

[9] John M. O'Toole (2022). Mutual information (`https://github.com/otoolej/mutual_info_kNN/releases/tag/v0.1.1`), GitHub. Retrieved December, 2022.